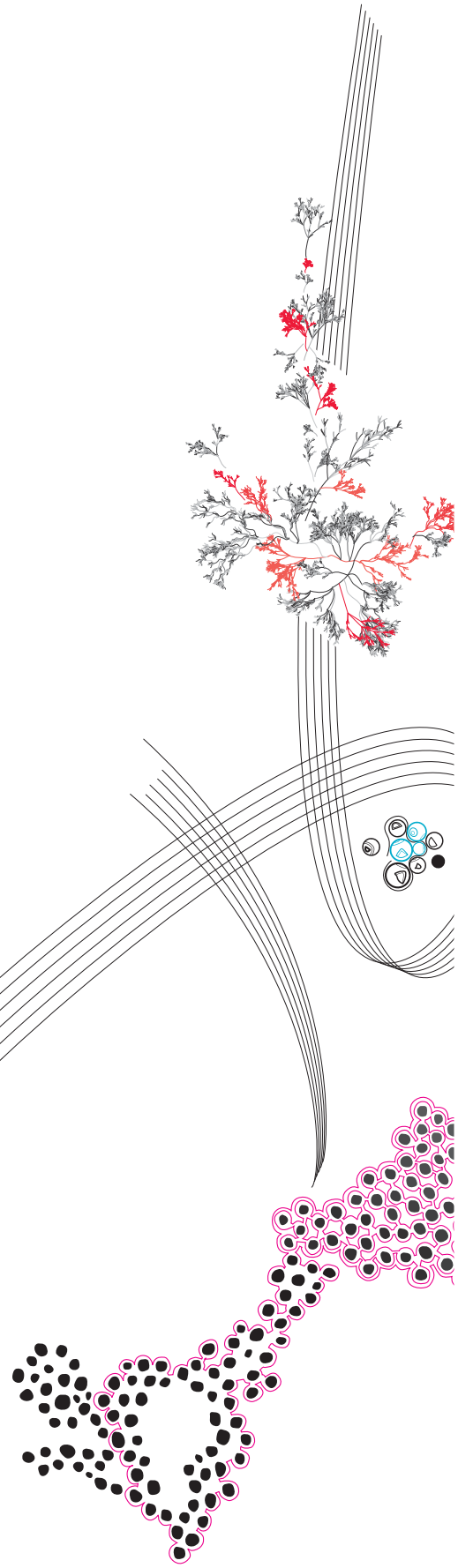


BSc Thesis Applied Mathematics

# Smoothed Analysis of the ICP Algorithm for Arbitrary Probability Distributions

C. Schmit

Supervisor: Dr. B. Manthey, Dr. P.T. De Boer

January, 2020

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

## Preface

This Bachelor thesis marks the end of my double degree Bachelor in Applied Mathematics and Technical Computer Science. I wish to express my sincere appreciation to my supervisors who continuously supported me during last half year: Dr. B. Manthey for making sure I was Mathematically precise and helping me to grasp new theorems and Dr. P.T. De Boer for providing me with new ideas and making sure my explanations were sound and understandable. I also want to express my gratitude to my fellow students Mariya Karlashchuk and Evelien van der Meer for peer-reviewing my draft paper. Last but not least, I want to thank my close friends and family for supporting me during my entire Bachelor program.

BSc Thesis Applied Mathematics

# Smoothed Analysis of the ICP Algorithm for Arbitrary Probability Distributions

C. Schmit *

January, 2020

**Abstract**

The Iterative Closest Point (ICP) algorithm is a commonly used algorithm to align two point clouds. Worst-case analysis gives an exponential upper bound on the running time, but the algorithm is observed to work more efficiently in practice. To reconcile this gap, Arthur and Vassilvitskii perform a smoothed analysis of the ICP by perturbing the inputs according to a Gaussian distribution. Their polynomial smoothed complexity implies that the ICP algorithm should perform well in practice. This paper generalizes their findings, by proving a polynomial smoothed complexity for two more general perturbation models. In both models, the points are drawn according to arbitrary probability density functions upper bounded by a perturbation parameter. In the first model, the functions' support is bounded to the unit hypercube. In the second model, their support is unbounded, but the probability distribution has exponentially decreasing tails.

*Keywords*: ICP algorithm, smoothed analysis, one-step model, perturbation model

## 1 Introduction

How to transform one set of points to optimally align it with a second point cloud? For example, this problem occurs regularly during rigid registrations in image analysis such as calibration of multiple cameras or the registration of 3D reconstructions [9]. The Iterative Closest Point (ICP) algorithm is commonly used to resolve this problem, as it is observed to be fast and to produce suitable results. However, despite its observed efficiency, theoretical analyses of its running time so far only arrived at a poor, exponential upper bound on the number of iterations [8].

To reconcile this gap between the practical and theoretical performance of an algorithm, Spielman and Teng developed 'smoothed analysis' [13]. Smoothed analysis is based on the idea that the worst-case instances of such algorithms come from a few specific instances that only occur rarely in practice. Therefore, the traditional worst-case analysis is too pessimistic and a different theoretical analysis is needed to back up the observations that the algorithm works well in everyday use cases. Therefore, instead of determining the worst-case running time over all instances, they first perturb the inputs by adding some Gaussian noise. Then, they determine the expected upper bound on the algorithm's number of iterations on these perturbed inputs. According to Spielman and Teng, algorithms with such a favorable smoothed complexity are expected to work well in practice. A reason for this is that practical inputs are often subjected to a certain degree of noise.

*Email: c.schmit@student.utwente.nl

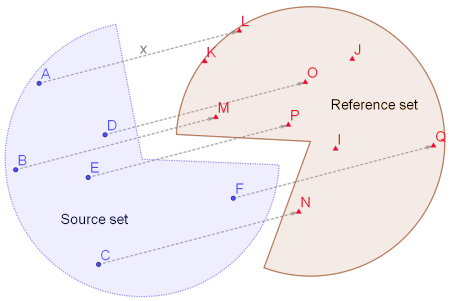



FIGURE 1: The ICP algorithm aims to find the translation $x$ that matches the source set to the reference set.

Generalizing this method, Beier and Vöcking proposed a different perturbation model, called the one-step model [2]. Instead of simply perturbing the input by Gaussian distributions, they perturb it according to arbitrary probability distributions. Their one-step model is "allowing the adversary to not just choose the mean of each input parameter, but even its distribution", as explained by T. Brunsch et al. [5].

Following the proof of the traditional smoothed analysis of the ICP algorithm [1], a smoothed analysis of the ICP algorithm for arbitrary probability distributions is performed using two different perturbation models. First, Section 2 provides a more detailed introduction to the ICP algorithm and into smoothed analysis which is needed for the following sections. In Section 3, a smoothed analysis is performed when the inputs are contained in the unit hypercube. To relax this condition, in Section 4, the perturbation model is based on arbitrary probability density functions with unbounded support so the points can be situated anywhere in the space. However, it should be noted that the latter scenario comes with an additional restriction. To be specific, the probability that points are far away from the origin decreases exponentially.

# 2 Preliminaries

Appendix A offers an overview of commonly used notations.

## 2.1 Iterative Closest Point algorithm (ICP)

**Introduction.** Algorithms such as the ICP algorithms were developed to solve the following optimization problem: How does a first set, the 'source set', need to be transformed to match a reference set, minimizing the distances between them. Besl and Mcay introduced the ICP algorithm matching the point clouds using translation and rotation [3]. Even though the ICP algorithm is mainly used in rigid registrations (so using translation or rotation), nowadays there are also variants to solve non-rigid registration problems in which the distance between the points could alter. For example, Du et al. proposed 'SICP', an ICP algorithm that includes scaling [7]. In addition to varying transformations, studies have also examined the efficiency of the ICP algorithm under different distance metrics [11].

**Analysed variant of the ICP algorithm.** The ICP algorithm analysed in this paper only allows translation and minimizes the least-squares measure according to the

Euclidean distance metric (see Figure 1). We refer to this objective function as 'potential function'. This variant is equal to the algorithm analysed by Arthur and Vassilvitskii [3]. Formally, we employ the following notation:

Let $d \in \mathbb{N}$. Given two finite point sets $A$ and $B$ in $\mathbb{R}^d$, the ICP algorithm seeks a translation $x \in \mathbb{R}$ and a nearest neighbour function $N_B : A \to B$, such that the potential $\Phi = \sum_{a \in A} \|a + x - N_B(a)\|^2$ is minimized. We refer to $d$ as dimension, $A$ as source set and $B$ as reference set. Note that $N_B(a)$ is the closest neighbour of the transformed $a$, so after being transformed by the current translation $x$.

**How does the algorithm work?** To minimize the potential, the algorithm employs two different strategies. Either the translation is fixed and the potential is minimized by updating the function $N_B$ or the nearest neighbouring function $N_B$ is fixed and the translation $x$ is optimized. In the first case, this means that the translated points in $A$ are reassigned to their closest neighbour in $B$. Note that multiple points in $A$ can be assigned to the same point in $B$. Secondly, having a fixed nearest neighbour assignment from all points in $A$ to points in $B$, the potential is minimized by adapting the translation $x$. The following proposition shows that that the potential is minimized when choosing $x = \frac{1}{|A|} \sum_{a \in A} (N_B(a) - a)$. The proposition is proved in Appendix A under Proposition B.2.

**Proposition 1.** *For a fixed $N_B : A \to B$, the potential $\Phi$ is minimized by choosing the translation*

$$x = \frac{1}{|A|} \sum_{a \in A} (N_B(a) - a).$$

Proposition 1 and the previous considerations explain the pseudo-code of the ICP algorithm, shown in Algorithm 1. We consider an *iteration* to be the consecutive execution of steps 3 up to 6.

**Previous theoretical results** An important tool in analysing the efficiency of an algorithm is the worst-case analysis. This has first been studied in detail by Ezra, Sharir and Efrat establishing the following upper bound on the number of iterations [8, Corollary 2.2].

**Theorem 2.** *In the worst case instance, the ICP algorithm needs at most $c(|A||B|d)^d$ iterations to terminate for some constant $c \in \mathbb{R}$.*

Arthur and Vassilvitskii prove the lower bound on the worst-case running time to be less than $\Omega(\frac{n}{d})^{d+1}$ [1]. To achieve this bound, they had to precisely construct a working example. However, the practice has shown that the algorithms generally work better than these bounds suggest [8, 12]. An explanation could be that these worst-case instances come from a few constructed instances that probably only rarely occur in practice.

## 2.2 Smoothed analysis

**Introduction** To study the efficiency of an algorithm, generally, a worst-case analysis is carried out first. However, in some cases, this worst-case analysis is too pessimistic.

**Algorithm 1** ICP algorithm

**Input**: $A \subseteq \mathbb{R}^d$: source point set, $B \subseteq \mathbb{R}^d$: reference point set
**Output**: Translation $x$ and nearest neighbour function $N_B$ minimizing the potential $\Phi$.

1: choose an initial translation $x$
2: **while** at least $x$ or $N_B$ (or both) is updated **do**
3: **for** each $a \in A$ **do**
4: $N_B(a) = \underset{b \in B}{\operatorname{argmin}} \|a + x - b\|$
5: **end for**
6: $x = \frac{1}{|A|} \sum_{a \in A} (N_B(a) - a)$
7: **end while**

For example, the ICP algorithm has an exponential worst-case complexity (Theorem 2), but it is observed to converge fast in practice and thus well-used [8]. To back up these observations, Spielman and Teng developed 'smoothed analysis' [13]. Spielman and Teng introduced smoothed analysis while studying the Simplex method: an algorithm with exponential worst-case complexity that often runs quicker than many other algorithms with polynomial worst-case complexity. Their convictions are that, probably, such worst-case instances rarely appear in practice and even a slight perturbation of the instance already results in shorter running time.

**Smoothed analysis using the two-step model** Assume the adversary in this train of thought to aim for the worst possible running time. Instead of considering every possible input, Spielman and Teng first let their adversary choose an instance and the perturbation parameter $\sigma$ [14]. Subsequently, this instance is perturbed by adding random noise according to a Gaussian distribution with standard deviation $\sigma$. Only then, the running time of the perturbed instance is examined. In short, the smoothed complexity is the maximal running time of all perturbed instances with respect to the perturbation parameter $\sigma$. With this in mind, it is clear why this perturbation model is called "two-step" model: the perturbation happens in two steps, first the instance, then the additional noise.

Smoothed analysis is a hybrid between worst-case analysis and average-case analysis. In case the perturbation is small, more specific instances can be pinpointed and consequently the analysis converges towards the worst-case analysis. On the other hand, if the perturbation parameter is large, the random noise is so significant that it becomes an average-case analysis with random inputs.

**Smoothed analysis using the one-step model** Generalizing the "traditional" smoothed analysis, Beier and Vöcking grant more power to the adversary [2]. In the one-step model, the adversary chooses the probability distribution according to which every input is drawn. This is not possible in the two-step model, in which the adversary can only control the standard deviation of the Gaussian noise added to his picked input. The smoothed complexity using the one-step model is in term of the perturbation parameter $\phi$. This parameter is the only restriction to the probability distribution chosen by the adversary: the probability density function must be bounded from above by $\phi$. This is why, during a smoothed analysis with the one-step model, no property inherited by certain probability distributions can be assumed.

The one-step model generalizes the two-step model because adding Gaussian noise

to a specific input is a feasible distribution if the perturbation parameter $\phi$ is larger than the maximum of the Gaussian distribution.

Similarly to the two-step model, the adversary's power lies in the perturbation parameter $\phi$. If $\phi = \infty$, it is possible to choose a distribution that only allows a specific instance (using the Dirac-Delta function). Consequently, the analysis becomes the worst-case analysis. Contrarily, for lower $\phi$, the inputs are more randomly distributed which is similar to an average-case analysis.

In Section 3, the smoothed analysis of the ICP algorithm is proved, using the one-step model as described by Curticapean and Künnemann [6]. They limit the domain of the probability density functions to the unit hypercube. For this reason, we have the following arbitrary probability density functions according to which the elements of $A$ and $B$ are drawn:

$$f : [0,1]^d \to [1, \phi]$$

To clarify what a feasible probability density function is, consider $f : [0,1]^d \to [1, \phi]$, when $\phi = 1$. Then, the uniform distribution is the only continuous distribution that satisfies $\int_{-\infty}^{\infty} f(x)\,dx = 1$ and $f(x) \leq \phi$, for all $x \in [0,1]^d$. The larger $\phi$, the more specific cases can be chosen (Figure 3). Another example of acceptable probability density functions is shown in Figure 2 when $\phi = 3$.

In Section 4, the perturbation model of the smoothed analysis does not restrict the domain of the probability density function.

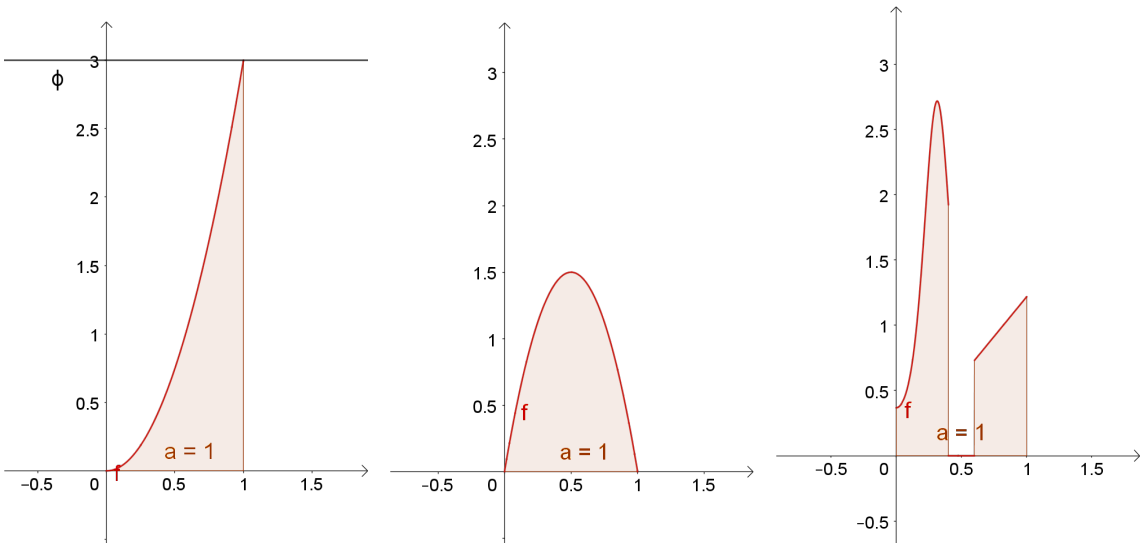



FIGURE 2: These figures show possible probability density functions $f : [0,1]^2 \to [0, \phi]$, where $\phi = 3$. To qualify, the functions must satisfy that $0 \leq f(x) \leq 3$, for all $x \in [0,1]^d$, and $a = \int_0^1 f(x)\,dx = 1$.

## 3 Generalized smoothed analysis of the ICP algorithm with bounded support

This section is dedicated to the performance of the smoothed analysis of the ICP algorithm, assuming that all points in $A$ and $B$ reside in the unit hypercube. The proof structure

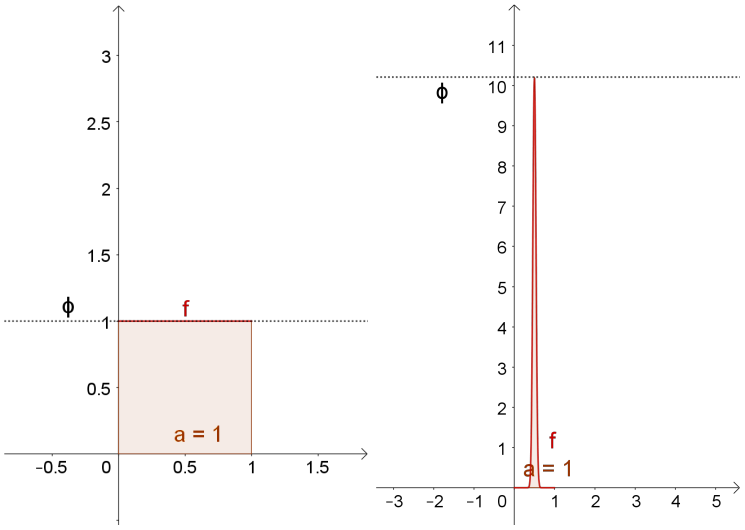



FIGURE 3: If $\phi = 1$, one of the few possible distribution is the uniform distribution, so the analysis is equivalent to the average-case analysis. In contrast, when $\phi$ is large, an adversary can pinpoint specific instances better.

follows the proof by Arthur and Vassilvjtskii who proved a polynomial upper bound for the ICP algorithm when the inputs are perturbed according to Gaussian distributions [1]. Some of their lemmas, revisit their proofs and adapt other lemmas to prove Theorem 3.

The following notation is used for all lemmas and propositions in the remainder of the section, except if specified otherwise. Some of the notation is also reused in Section 4. Let $d \in \mathbb{N}$. Let $A \subseteq \mathbb{R}^d$ and $B \subseteq \mathbb{R}^d$ be the source and the reference set, respectively. Let $\phi \in [1, \infty)$ be the perturbation parameter. Define $n = |A|$ and $m = |B|$. Let $f_i : [0,1]^d \to [1, \phi]$ for all $i = 1, 2, ..., n$ and $g_j : [0,1]^d \to [1, \phi]$ for all $j = 1, 2, ..., m$ be probability density functions. Define the functions $f$ and $g$ by $f = (f_1, f_2, ..., f_n)$ and $g = (g_{1,2}, ..., g_n)$, respectively. Let the elements of $A$ and $B$ be drawn according to these probability density functions. So $A = \{a_1, a_2, ..., a_n\}$ and $B = \{b_1, b_2, ..., b_m\}$ with $a_i \sim f_i$ for $i \in 1, 2, ..., n$ and $b_j \sim g_j$ for $j \in 1, 2, \ldots, m$. Furthermore, $a_{ik}$, refers to the $k$th coordinate of the $i$th element in $A$.

The ICP algorithm is performed on $A$ and $B$ to determine a translation $x$ and a nearest neighbouring function $N_B : A \to B$ that (locally) minimizes the potential $\Phi = \sum_{a \in A} \|a + x - N_B(a)\|^2 = \sum_{i=1}^{n} \|a_i + x - N_B(a_i)\|^2$.

**Theorem 3** (Generalized smoothed complexity for the ICP algorithm, bounded support)**.** *Let the elements of $A$ and $B$ be defined as described above. Then, the expected number of iterations of the ICP algorithm is bounded from above by a polynomial depending on $n$, $m$, $d$ and $\phi$.*

To find an upper bound for the expected amount of iterations, the following approach is applied:

1. First, we prove that the potential $\Phi$ has a polynomial upper bound, denoted by $\Phi_0$, depending only on $n$, $m$, $\phi$ and $d$.

2. Let $\Delta$ be the decrease of the potential during an iteration after the first iteration. We prove that with high probability the decrease of the potential $\Phi$ per iteration is larger or equal to a strictly positive number, called $\Delta_0$.

3. Let $W$ denote the worst-case upper bound on the number of iterations (Theorem 2). Define the number of iterations by the random variable $T$. Then, the expectation of $T$ is bounded by above as follows:

$$\begin{aligned} E(T) &= E(T \mid \Delta < \Delta_0) \cdot P(\Delta < \Delta_0) + E(T \mid \Delta \geq \Delta_0) \cdot P(\Delta \geq \Delta_0) \\ &\leq W \cdot P(\Delta < \Delta_0) + E(T \mid \Delta \geq \Delta_0) \\ &\leq W \cdot P(\Delta < \Delta_0) + \frac{\Phi_0}{\Delta_0}. \end{aligned}$$

The last line follows from the fact that the potential $\Phi$ cannot be negative. This implies that the algorithm needs to terminate after at most $\frac{\Phi_0}{\Delta_0}$ iterations. Otherwise, after an additional iteration, $\Phi - \left(\frac{\Phi_0}{\Delta_0} + 1\right)\Delta_0 \leq \Phi_0 - \left(\frac{\Phi_0}{\Delta_0} + 1\right)\Delta_0 = -\Delta_0 < 0$.

We are going to show that $P(\Delta < \Delta_0)$ is small enough such that $W \cdot P(\Delta < \Delta_0)$ is polynomial. In addition to this, it is proven that $\frac{\Phi_0}{\Delta_0}$ is polynomial. Consequently, it is shown that the expected number of iterations is bounded from above by a polynomial (in $n$, $m$, $d$, and $\phi$).

### 3.1 Upper bound of the potential

To find an upper bound on the potential, the first iteration of the algorithm is ignored, as there is no information about the initial translation. First, the domain of the translation $x$ after one iteration is identified. Subsequently, this allows an upper bound on the squared distance of two points $a \in A$ and $b \in B$ to be identified. Finally, this result is used to determine an upper bound for the potential.

**Lemma 4.** *Let $A \subseteq [0,1]^d$ and $B \subseteq [0,1]^d$. After every iteration of the ICP algorithm, the translation $x$ is in $[-1,1]^d$.*

*Proof.* After every iteration, the nearest neighbour function $N_B$ assigns every element in $A \subseteq [0,1]^d$ to some element in $B \subseteq [0,1]^d$. Let $a_i$ be the $i$th element in $A$ and denote $N_B(a_i)$ by $b_i$, so $b_i \in B$. Then, the one-dimensional components $a_{ik} \in [0,1]$ and $b_{ik} \in [0,1]$, for all $k \in \{1, \ldots, d\}$. Furthermore, based on the current $N_B$, we obtain the new translation $x = \frac{1}{n}\sum_{i=1}^{n}(b_i - a_i)$. Then for all $k \in \{1, \ldots d\}$,

$$|x_k| = \left|\frac{1}{n}\sum_{i=1}^{n}(b_{ik} - a_{ik})\right| \leq \frac{1}{n}\sum_{i=1}^{n}|b_{ik} - a_{ik}| \leq \frac{1}{n}\sum_{i=1}^{n}1 = \frac{n}{n} = 1.$$

Thus, $x_k \in [-1,1]$. Hence, $x \in [-1,1]^d$. □

**Lemma 5.** *Let $a_i \in A$ be fixed. Let $N_B$ be the nearest neighbour function and $x$ be the translation $x$ after the first iteration of the ICP algorithm. Then,*

$$\|a_i + x - N_B(a_i)\|^2 \leq 4d,$$

*for all $i \in \{1, 2, \ldots, n\}$.*

*Proof.* All elements of $A$ and $B$ are chosen from probability density functions $f_i : [0,1]^d \to [0, \phi]$ and $g_j : [0,1]^d \to [0, \phi]$, $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, respectively. We know that $A \subseteq [0,1]^d$ and $B \subseteq [0,1]^d$. According to Lemma 4, the translation $x \in [-1,1]^d$ after the first iteration of the ICP algorithm. For all $a_i \in A$, this implies that $a_i + x \in [-1,2]^d$, since $a_{ik} + x_k \leq 1 + 1 = 2$ and $a_{ik} + x_k \geq 0 - 1 = -1$ for $k \in \{1, 2, \ldots, d\}$.

After the first iteration, $N_B$ assigns each element in $A$ to an element in $B$. So we know that, $N_B(a_i) \in [0,1]^d$ and $a_i + x \in [-1,2]^d$, for all $a_i \in A$. This implies that $a_i + x - N_B(a_i) \in [-2,2]^d$. Therefore,

$$\|a_i + x - N_B(a_i)\|^2 \leq 2^2 + 2^2 + \cdots + 2^2 = 4d,$$

for all $i \in \{1, 2, \ldots, n\}$. □

Using Lemma 5 the upper bound of the potential $\Phi$ can be determined.

**Proposition 6** (Upper bound on the potential)**.** *Let $N_B$ be the nearest neighbour function and $x$ be the translation $x$ after the first iteration of the ICP algorithm. Then,*

$$\Phi = \sum_{i=1}^{n} \|a_i + x - N_B(a_i)\|^2 \leq 4nd.$$

*Proof.* Lemma 5 states that $\|a_i + x - N_B(a_i)\|^2 \leq 4d$ for every $a_i \in A$, so

$$\Phi = \sum_{i=1}^{n} \|a_i + x - N_B(a_i)\|^2 \leq \sum_{i=1}^{n} 4d = 4dn.$$

□

The upper bound for the potential $\Phi$ is denoted by $\Phi_0 = 4dn$.

After establishing an upper bound for the potential, we study the prospective drop of the potential after every iteration.

### 3.2 Decrease of the potential function per iteration

In this section, the decrease of the potential function during an arbitrary iteration of the ICP algorithm (after the first iteration) is studied. In general $z_{(1)}$ is considered to be the expression of the variable $z$ before this iteration, while $z_{(2)}$ is the value after this iteration. To find the decrease of the potential, two cases are considered depending on whether more or less than $k$ points of $A$ are reassigned to new points in $B$ during this iteration. The exact expression of $k$ is to be determined later. An important result to remember from Subsection 2.1 is that a decrease in the potential $\Phi$ is either due to reassignments (so a change of $N_B$) or due to a new translation $x$. We first explain the proof outline for both cases and then proceed with the details.

- **There are at most $k$ reassignments**. This means that there are at most $k$ distinct points in $A$ for which the nearest neighbour function $N_B$ changes. As there is a low amount of reassignments, a substantial decrease in the potential due to changes in $N_B$ might be unlikely, so we concentrate on the drop resulting from a new translation $x$. We prove with a high probability that the points in $B$ are sparse, so scattered and not condensed in a single area. This property ensures a significant drop of the potential during one iteration.

- **There are at least $k$ reassignments**. We prove that if a point $a \in A$ is not approximately equidistant to the former closest point $b_{(1)} \in B$ and the new closest point $b_{(2)} \in B$, then many reassignments results in a significant drop of the potential. Furthermore, this condition is proven to hold with high probability.

### 3.2.1 At most $k$ reassignments

In this section, the drop resulting from the change in translation and the fact that the points in $B$ are 'sparse' is analysed. Therefore, the next lemma is useful.

**Lemma 7.** *Let $x_{(1)}$ and $x_{(2)}$ be the translation before and after an iteration of the ICP algorithm, respectively. Then, the potential $\Phi$ decreases by at least $n \cdot \left\|x_{(1)} - x_{(2)}\right\|^2$ during this iteration.*

*Proof.* According to Proposition 1, the potential is minimized by choosing the center of mass as translation, $x = \frac{1}{|A|}\sum_{a\in A}(N_B(a) - a)$. Let $S = \{N_B(a) - a \mid a \in A\}$. Furthermore, as derived in Lemma B.1, the potential can be written as

$$\Phi = \sum_{a\in A} \|a + x - N(a)\|^2 = \sum_{s\in S} \|x - s\|^2 = \sum_{s\in S} \|s - c(S)\|^2 + |S| \cdot \|c(S) - x\|^2 .$$

Now, the potential is examined at two different points during the iteration.
After updating $N_B$ and before changing the translation from $x_{(1)}$ to $x_{(2)}$ (between steps 5 and 6 in Algorithm 1), the potential $\Phi_1$ is as follows,

$$\Phi_1 = \sum_{s\in S} \|s - c(S)\|^2 + |S| \cdot \left\|c(S) - x_{(1)}\right\|^2 . \tag{1}$$

After updating the translation (executing line 6 in Algorithm1), the translation changes to $x_{(2)} = c(S)$. Then, the potential changes to

$$\Phi_2 = \sum_{s\in S} \|s - c(S)\|^2 + |S| \cdot \left\|c(S) - x_{(2)}\right\|^2 = \sum_{s\in S} \|s - c(S)\|^2 . \tag{2}$$

Combining Equations (1) and (2), the following drop of the potential is experienced during an iteration. Note that $|S| = |A| = n$:

$$\begin{aligned}\Phi_1 - \Phi_2 &= \sum_{s\in S} \|s - c(S)\|^2 + |S| \cdot \left\|c(S) - x_{(1)}\right\|^2 - \sum_{s\in S} \|s - c(S)\|^2 \\ &= |S| \cdot \left\|c(S) - x_{(1)}\right\|^2 = n \left\|x_{(2)} - x_{(1)}\right\|^2 .\end{aligned}$$

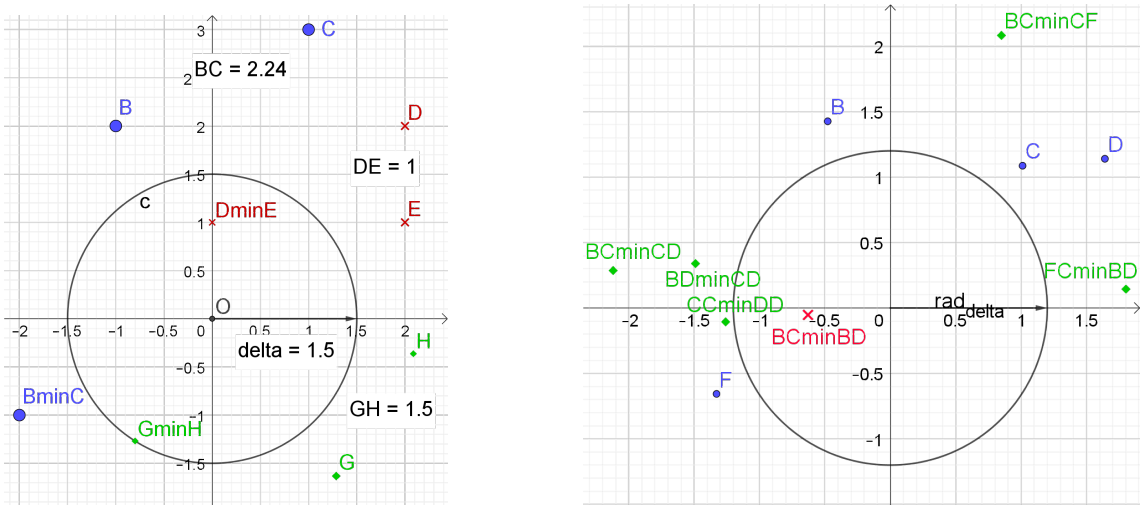
□

Before continuing, we clarify what is meant with 'the elements in $B$ are sparse'.

**Definition 8** ($(k,\delta)$-sparse)**.** A set $B$ is $(k,\delta)$-sparse if no pair of distinct size-$k$ multisets[1] $B_1, B_2 \subset B$ satisfies $\left\|\sum_{b_1\in B_1} b_1 - \sum_{b_2\in B_2} b_2\right\| \leq \delta$ (cf. Arthur and Vassilvitskii [1, Definition 4.2]).

To get an intuitive understanding of what this means, consider the definition when $k = 1$. Then, the definition states that $B$ is $(1,\delta)$-sparse if no pair of distinct elements $b_1, b_2 \in B$ satisfies $\|b_1 - b_2\| \leq \delta$. This means that all elements are separated by at least a distance $\delta$ from every other element (Figure 4a). A different illustration of the definition is given in Figure 4b.
Combining the previous lemma and definition, a significant drop of the potential per iteration can be determined.

[1] See Appendix A.3




(A) $d = 2$, $k = 1$, $\delta = 1.5$ (B) $d = 2$, $k = 2$, $\delta = 1.2$

FIGURE 4: Let '$D\text{min}E$' denote the point calculated by $D - E$ and '$BC\text{min}BD$' the point calculated by $B + C - (B + D)$.
On the left: the set $\{B, C, D, E, H, G\}$ is not $(k = 1, \delta = 1.5)$-sparse, as the difference between $E$ and $D$ is less than 1.5. This can be seen by the fact that '$D\text{min}E$' lies within the circle of radius 1.5.
On the right: the set $\{B, C, D, E\}$ is not $(k = 2, \delta = 1.2)$-sparse, as the difference between $B + C$ and $B + D$ is less than 1.2.

**Proposition 9.** *Let $A, B \subseteq \mathbb{R}^d$ and let $B$ be $(k, \delta)$-sparse. Suppose that during an iteration of the ICP algorithm, after the first iteration, there are at most $k$ reassignments. Then, the potential $\Phi$ decreases by at least $\frac{\delta^2}{n}$ or the ICP algorithm terminates (cf. Arthur and Vassilvitskii [1, Proposition 4.3]).*

*Proof.* Let $x_{(1)}$ and $x_{(2)}$ be the translation before and after an iteration. Similarly, define $N_{B(1)}$ and $N_{B(2)}$. Let $j \in \{1, 2\}$. Then,

$$x_{(j)} = \frac{1}{n} \sum_{a \in A} \left( N_{B(j)}(a) - a \right) = \frac{1}{n} \left( \sum_{a \in A} N_{B(j)}(a) - \sum_{a \in A} a \right).$$

Consequently,

$$\begin{aligned} \left\| x_{(1)} - x_{(2)} \right\| &= \left\| \frac{1}{n} \left( \sum_{a \in A} N_{B(1)}(a) - \sum_{a \in A} a \right) - \frac{1}{n} \left( \sum_{a \in A} N_{B(2)}(a) - \sum_{a \in A} a \right) \right\| && (3) \\ &= \frac{1}{n} \left\| \sum_{a \in A} N_{B(1)}(a) - \sum_{a \in A} a - \sum_{a \in A} N_{B(2)}(a) + \sum_{a \in A} a \right\| && (4) \\ &= \frac{1}{n} \left\| \sum_{a \in A} \left( N_{B(1)}(a) - N_{B(2)}(a) \right) \right\|. && (5) \end{aligned}$$

As there are at most $k$ reassignments, let the multiset [1] $B_j = \{N_{B(j)}(a) \mid \forall a \in A : N_{B(1)}(a) \neq N_{B(2)}(a)\} \subset B$ be the set of all images according to $N_{B(j)}$ of these (at most)

[1] See Appendix A.3

$k$ points in $A$. For all other points $a^* \in A$, it holds that $N_{B(1)}(a^*) = N_{B(2)}(a^*)$, so they cancel each other out in Equation (5). Thus,

$$\left\|x_{(1)} - x_{(2)}\right\| = \frac{1}{n}\left\|\sum_{b_1 \in B_1} b_1 - \sum_{b_2 \in B_2} b_2\right\|. \tag{6}$$

If $B_1$ and $B_2$ are identical, the potential is not reduced due to the reassignment and the translation remains unchanged, so the algorithm terminates.
If $B_1$ and $B_2$ are not identical, by definition $|B_1| = |B_2| \leq k$. Adding $s = k - |B_1|$ times the same element $b^* \in B$ to each set creates new multisets $B_1^*$ and $B_2^*$ of size $k$. Since $B$ is $(k, \delta)$-sparse, this implies that

$$\delta < \left\|\sum_{b_1 \in B_1^*} b_1 - \sum_{b_2 \in B_2^*} b_2\right\| = \left\|\sum_{b_1 \in B_1} b_1 - \sum_{b_2 \in B_2} b_2 + s \cdot (b^* - b^*)\right\| = \left\|\sum_{b_1 \in B_1} b_1 - \sum_{b_2 \in B_2} b_2\right\|.$$

Combining this with Equation (6),

$$\left\|x_{(1)} - x_{(2)}\right\| = \frac{1}{n}\left\|\sum_{b_1 \in B_1} b_1 - \sum_{b_2 \in B_2} b_2\right\| > \frac{\delta}{n}.$$

By Lemma 7, the potential decreases at least by $n\left(\frac{\delta}{n}\right)^2 = \frac{\delta^2}{n}$. □

What remains is to prove that $B$ is $(k, \delta)$-sparse with high probability. For this, an auxiliary lemma is proven first, before stating the final proposition.

**Lemma 10.** *Let $S \subseteq \mathbb{R}^d$ and let $h$ be a probability density function of the form, $h : S \to [0, \phi]$. Let $z$ be chosen according to $h$. Then the probability that $z$ is in a fixed ball of radius $r$ is at most $\phi(2r)^d$ (cf. Curticapean and Künnemann [6, Chapter 2]).*

*Proof.* In every dimension, a $d$-dimensional ball $B$ of radius $r$ is inscribed in a $d$-dimensional hypercube $C$ with side length $2r$. This implies that the volume of the hyperball is smaller than the volume of this hypercube which equals $(2r)^d$. Furthermore, the upper bound of the probability density function equals $\phi$. This gives

$$P(z \in B) = \int_B h(x)\,dx < \int_B \phi\,dx \leq \int_C \phi\,dx = \text{volume}(C)\phi = \phi(2r)^d.$$

□

**Proposition 11.** *The probability that $B$ is not $(k, \delta)$-sparse is at most $m^{2k}\phi(2\delta)^d$ (cf. Arthur and Vassilvitskii [1, Proposition 4.4]).*

*Proof.* Remember that $B$ is not $(k, \delta)$-sparse, if and only if, there exists a pair of two distinct size-$k$ multisets[1] $B_1, B_2 \subset B$, such that $\left\|\sum_{b_1 \in B_1} b_1 - \sum_{b_2 \in B_2} b_2\right\| \leq \delta$. First, the probability is determined that any two distinct size-$k$ multisets of $B$, named $C_1$ and $C_2$, satisfy $\left\|\sum_{c_1 \in C_1} c_1 - \sum_{c_2 \in C_2} c_2\right\| \leq \delta$. Afterwards the proposition is proven using the union bound (Lemma B.1) and the fact that there are only $|B|^k = m^k$ possible k-sized multisets of $B$.

[1] See Appendix A.3

Let $C_1$ and $C_2$ be two size-$k$ multisets such that $C_1, C_2 \subset B$ and $C_1 \neq C_2$. As $C_1 \neq C_2$, there exists an element $c^*$ such that its multiplicity in $C_1$ is strictly larger than its multiplicity in $C_2$. Let all other elements in $C_1$ and $C_2$ be fixed arbitrarily, while $c^*$ is drawn according to the distribution defined for the elements of $B$. Let $\ell \geq 1$ be the multiplicity of $c^*$ in $C_1 - C_2$. Then $\left\|\sum_{c_1 \in C_1} c_1 - \sum_{c_2 \in C_2} c_2\right\|$ can be written as $\|c \pm \ell c^*\|$, where $c$ is the constant resulting from the sum and difference of all other fixed elements in $C_1$ and $C_2$. Then,

$$\left\|\sum_{c_1 \in C_1} c_1 - \sum_{c_2 \in C_2} c_2\right\| = \|c \pm \ell c^*\| \leq \delta \iff \left\|\frac{c}{\ell} \pm c^*\right\| \leq \frac{\delta}{\ell}$$

The event that $\left\|\frac{c}{\ell} \pm c^*\right\| \leq \frac{\delta}{\ell}$ is equal to the event that $c^*$ lies in a ball of radius $\frac{\delta}{\ell}$ (centered at $\mp\frac{c}{\ell}$). By Lemma 10 this event happens with a probability of at most $\phi(2\frac{\delta}{\ell})^d \leq \phi(2\delta)^d$.

As there are $m^k$ choices for $C_1$ and (less than) $m^k$ choices for $C_2$, there are less than $m^{2k}$ choices for a combination of $C_1$ and $C_2$. By the union bound,

$$\begin{aligned}
&P(B \text{ is not } (k,\delta)\text{-sparse}) \\
&= P\left(\exists B_1, B_2 \subset B \text{ with } B_1 \neq B_2 \text{ s.t. } \left\|\sum_{b_1 \in B_1} b_1 - \sum_{b_2 \in B_2} b_2\right\| \leq \delta\right) \\
&\leq \sum_{i=1}^{m^{2k}} P\left(\left\|\sum_{c_1 \in C_1} c_1 - \sum_{c_2 \in C_2} c_2\right\| \leq \delta\right) \leq m^{2k}\phi(2\delta)^d.
\end{aligned}$$

□

The event that $B$ is not $(k,\delta)$-sparse (Proposition 11) is called 'Failure 1' and denoted by $F_1$.

This completes the analysis of the decrease in potential during an iteration when at most $k$ elements are reassigned: with probability at least $1 - P(F_1)$, the potential decreases at least by $\frac{\delta^2}{n}$.

### 3.2.2 At least $k$ reassignments

In this section, the drop resulting when there are at least $k$ reassignments is analysed. First, some definitions are introduced.

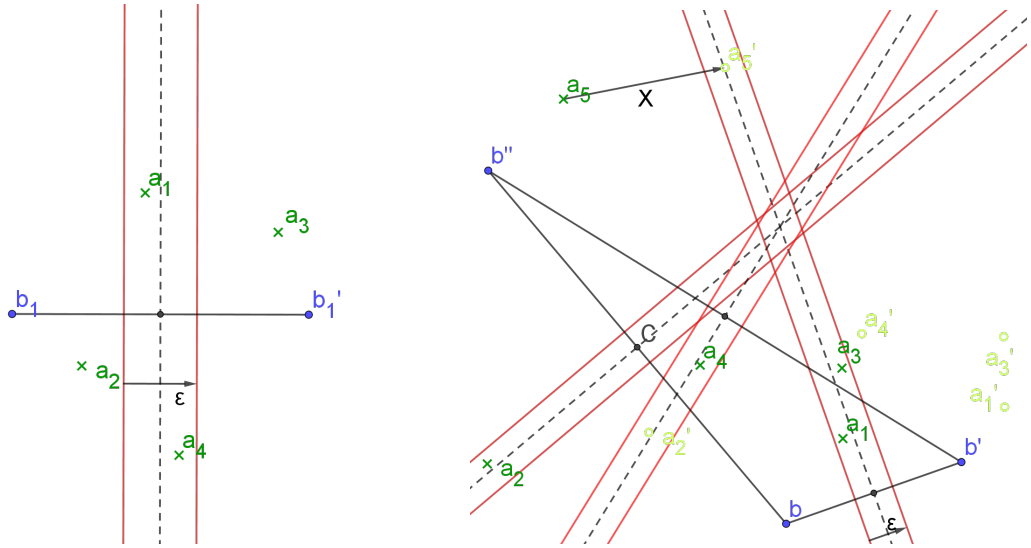
**Definition 12** ($\epsilon$-centered)**.** Let $a, b, b' \in \mathbb{R}^d$. The element $a$ is said to be $\epsilon$-centered between $b$ and $b'$ if $a$ is within a distance $\frac{\epsilon}{2}$ of the hyperplane[2] bisecting $b$ and $b'$ (cf. Arthur and Vassilvitskii [1, Definition 4.5]).

**Definition 13** ($(k,\epsilon)$-centerable)**.** Let $A, B \subseteq \mathbb{R}^d$. $(A, B)$ is said to be $(k,\epsilon)$-centerable if there exist distinct $a_1, a_2, \ldots, a_k \in A$, not necessarily distinct $b_1, b_2, \ldots, b_k, b'_1, b'_2, \ldots, b'_k \in B$ and a translation $x \in \mathbb{R}^d$ such that $a_i + x$ is $\epsilon$-centered between $b_i$ and $b'_i$ for all $i = 1, 2, \ldots, k$ (cf. Arthur and Vassilvitskii [1, Definition 4.7]).

See Figure 5 for a better understanding of the above definitions.

**Proposition 14.** *Let $A, B \subseteq \mathbb{R}^d$ be finite point sets such that $(A, B)$ is not $(k,\epsilon)$-centerable. Consider an iteration of the ICP algorithm after the first one. If the nearest neighbour function $N_B$ changes its value for at least $k$ points during this iteration, it decreases the potential $\Phi$ by at least $n\frac{\epsilon^2}{4}$ (cf. Arthur and Vassilvitskii [1, Proposition 4.8]).*

[2] See Appendix A.4




(A) $d = 2$, which points are $\epsilon$-centered? (B) $d = 2$, is $(A_1, B_1)$ $(k, \epsilon)$-centerable?

FIGURE 5: On the left: the points $a_1$ and $a_4$ are $\epsilon$-centered between $b_1$ and $b_1'$ while $a_2$ and $a_3$ are not.
On the right: $A_1 = \{a_1, a_2, a_3, a_4, a_5\}$ and $B_1 = \{b, b', b''\}$. $(A_1, B_1)$ is $(4, \epsilon)$-centerable, as there exist $A = \{a_1, a_2, a_3, a_4\}$ from $A_1$, $B = \{b, b, b, b'\}$ from $B_1$, $B' = \{b', b'', b', b_1''\}$ from $B_1$ and $x = 0$, such that $a_i + x$ is $\epsilon$-centered between $b_i$ and $b_i'$ for all $i = 1, 2, 3, 4$. However, $(A_1, B_1)$ is not $(5, \epsilon)$-centerable, as there is no translation that allows all points of A to be $\epsilon$-centered between elements of $B$. An example translation $x$ is shown with the translated points of $A$.

*Proof.* Let $a_1, a_2, \ldots, a_k$ be $k$ different points in $A$ for which $N_B$ changes. The previous assignment is denoted by $b_i = N_{B(1)}(a_i) \in B$ and the new assignment is written as $b_i' = N_{B(2)}(a_i) \in B$ for all $i = 1, 2, \ldots, k$. Furthermore, $x_{(1)}$ is the translation before this iteration while $x_{(2)}$ is the new translation.

Since $(A, B)$ is not $(k, \epsilon)$- centerable, there exists an $i \in \{1, \ldots, k\}$ such that $a_i + x_{(2)}$ is not $\epsilon$-centered between $b_i$ and $b_i'$. This means $a_i + x_{(2)}$ is not within a distance at most $\frac{\epsilon}{2}$ of the hyperplane bisecting $b_i$ and $b_i'$. Denote this hyperplane by $\mathcal{H}$.

A hyperplane splits $\mathbb{R}^d$ into two different half-spaces[3] in which "any line segment joining two points from different half-spaces must cut the hyperplane" [4]. As $\mathcal{H}$ bisects $b_i$ and $b_i'$, the space $\mathbb{R}^d$ is split into all points closer to $b_i$ in one half-space and all points closer to $b_i'$ in the other half-space. Furthermore, $a_i + x_{(1)}$ is closer to $b_i$ than to $b_i'$, otherwise $a_i$ would not have been assigned to the element $b_i$ in $B$. Similarly, $a_i + x_{(2)}$ is closer to $b_i'$ than to $b_i$. This implies that $a_i + x_{(2)}$ and $a_i + x_{(1)}$ are not in the same half-space and the line segment between them has to go through the hyperplane.

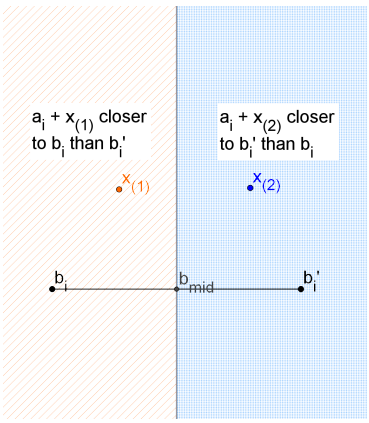
Combining this with the fact that $a_i + x_{(2)}$ is not within a distance of at most $\frac{\epsilon}{2}$ of $H$, we can draw the following conclusion about the distance between $a_i + x_{(1)}$ and $a_i + x_{(2)}$:

$$\left\|x_{(1)} - x_{(2)}\right\| = \left\|a_i + x_{(1)} - (a_i + x_{(2)})\right\| > \frac{\epsilon}{2}.$$

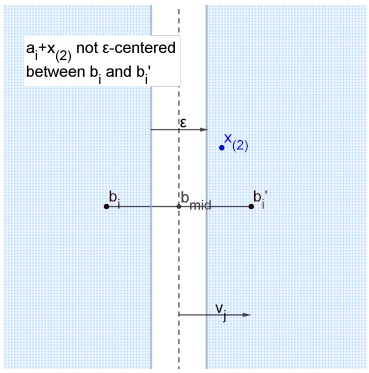
Thus, $\left\|x_{(1)} - x_{(2)}\right\| > \frac{\epsilon}{2}$ and by Lemma 7 the potential decreases by $n\left(\frac{\epsilon}{2}\right)^2$. □

[3] See Appendix A.5

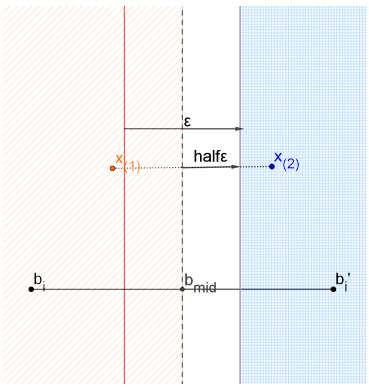
An illustration for Proposition 14 when $d = 2$ is explained in Figure 6.




(A) $a_i + x_{(1)}$ closer to $b_i$ than $b_i'$ and $a_i + x_{(2)}$ closer to $b_i'$ than $b_i$. The domains of $x_{(2)}$ and $x_{(1)}$ divide the plane.




(B) $a_i + x_{(2)}$ is not $\epsilon$-centered between $b_i$ and $b_i'$. This means $x_{(2)}$ must be outside the slab of width $\epsilon$ in the direction of $v_j$ centered at $b_{\text{mid}}$.




(C) Combining the conditions of figures 6a and 6b, it is clear that wherever $x_{(1)}$ and $x_{(2)}$ are defined in their domain, $\left\|x_{(1)} - x_{(2)}\right\| > \frac{\epsilon}{2}$.

FIGURE 6: This figure illustrates that $\left\|x_{(1)} - x_{(2)}\right\| > \frac{\epsilon}{2}$. Here, $a_i = 0$.

To complete this section, it must still be proven that $(A, B)$ is not $(k, \epsilon)$-centerable with a high probability. To prove this proposition, two more lemmas are needed. The first technical lemma is proven by Arthur and Vassilvitskii [1, Lemma 4.9].

**Lemma 15.** *Let $V$ denote a point set in $\mathbb{R}^d$. Then, there exists $V_0 \subset V$ with $|V_0| = d$ such that any $v \in V$ can be expressed as $\sum_{u \in V_0} c_u u$ for scalars $c_u \in [-1, 1]$ (cf. Arthur and Vassilvitskii [1, Lemma 4.9]).*

**Lemma 16.** *Let $z = (z_1, z_2, \ldots, z_d)$ be an element of $A$. Let $v = (v_1, v_2, \ldots, v_d) \in \mathbb{R}^d$ be a unit vector. Let $I \subset \mathbb{R}$ be an interval of length $\epsilon$. Then, the probability that the dot product $z \cdot v$ lies in $I$ is at most $\epsilon\phi\sqrt{d}$,*

$$P(z \cdot v \in I) \leq \epsilon\phi\sqrt{d}.$$

*Proof.* Let $I = [c, e]$, with $c, e \in \mathbb{R}$ and $e - c = \epsilon$. Let $v_i$ be the component of $v$ with the largest absolute value, so $v_i = \text{argmax}_{v_* \in v}|v_*|$. By Lemma B.3, $|v_i| \geq \frac{1}{\sqrt{d}}$, implying that $\frac{1}{|v_i|} \leq \sqrt{d}$. Furthermore, if we fix all components of $x$ and $v$ except $x_i$ and $v_i$, then $z \cdot v = z_1 v_1 + z_2 v_2 + \cdots + z_d v_d = z_i v_i + C$, with some $C \in \mathbb{R}$. Furthermore,

$$z \cdot v \in I \iff c \leq z \cdot v \leq e \iff c \leq z_i v_i + C \leq e \iff z_i v_i \in [c - C, e - C].$$

Then, $z \cdot v \in I$ is equivalent to $z_i v_i \in [c^*, e^*]$, with $c^* = c - C$ and $e^* = e - C$. In other

words, that $z_i v_i$ lies in an interval of length $e^* - c^* = e - c = \epsilon$. So,

$$\begin{aligned}
P(z \cdot v \in I) \leq & \int_0^1 \int_0^1 \dots \int_0^1 \int_{\frac{1}{v_i}c^*}^{\frac{1}{v_i}e^*} f(z_1 = x_1, z_2 = x_2 \dots, z_d = x_d)\, dx_i\, dx_1 \dots\, dx_d \\
\leq & \int_0^1 \int_0^1 \dots \int_0^1 \int_{\frac{1}{v_i}c^*}^{\frac{1}{v_i}e^*} \phi\, dx_i\, dx_1 \dots\, dx_d \\
= & \int_0^1 \int_0^1 \dots \int_0^1 \frac{1}{v_i} \left[e^* - c^*\right] \phi\, dx_1\, dx_2 \dots\, dx_d \\
= & \int_0^1 \int_0^1 \dots \int_0^1 \frac{1}{v_i} \epsilon\phi\, dx_1\, dx_2 \dots\, dx_d \\
= & 1^{d-1} \frac{1}{v_i} \epsilon\phi \leq \sqrt{d}\epsilon\phi.
\end{aligned}$$

□

**Proposition 17.** *Let $k \geq d$. Then $(A, B)$ is $(k, \epsilon)$-centerable with probability at most $(nm^2)^k \left((d+1)\epsilon\sqrt{d}\phi\right)^{k-d}$ (cf. Arthur and Vassilvitskii [1, Corollary 4.11]). In other terms,*

$$P(\text{$(A, B)$ is $(k, \epsilon)$-centerable}) < (nm^2)^k \left((d+1)\epsilon\sqrt{d}\phi\right)^{k-d}.$$

*Proof.* Let $a_1, a_2, \dots, a_d, \dots, a_k$ be $k$ arbitrary, distinct points in $A$ for which $N_B$ changes. The previous assignment of $a_1$ is denoted by $b_1 = N_{B(1)}(a_1) \in B$ and the new assignment is written as $b'_1 = N_{B(2)}(a_1) \in B$. The same notation is used for all other $k-1$ points. We only use the randomness in $a_1, a_2, \dots, a_d, \dots, a_k$, so all $b_1, b_2, \dots, b_d, \dots, b_k$ and $b'_1, b'_2, \dots, b'_d, b'_k$ are fixed arbitrarily, not necessarily distinctly.

Define a *slab* by its center and height $h$ in the direction of a vector $v$. A slab is a $d$-orthotope (generalized rectangle) with unbounded length in $d-1$ dimensions and bounded height in the direction of $v$. In addition to this, different index notations are used to distinguish between all $k$ points (index $\ell$), the first $d$ points (index $k$) and some point within the last $k - d$ points (index $j$).

For all $\ell \in \{1, 2, \dots, d, \dots, k\}$, let $v_\ell = \frac{b'_\ell - b_\ell}{\|b'_\ell - b_\ell\|}$ be the unit vector in $b'_\ell - b_\ell$ direction. Let $V = \{v_1, v_2, \dots, v_d, \dots v_k\}$. Let $U = \{v_1, v_2, \dots, v_d\}$ satisfy the properties of $V_0$ of Lemma 15 for $V$. This means that every element $v \in V$ can be written as $v = \sum_{u \in U} c_{u,v} u$ with coefficients $c_{u,v} \in [-1, 1]$. Let $X$ be the set of all translations $x \in \mathbb{R}$ that satisfy that $a_i + x$ is $\epsilon$-centered between $b_i$ and $b'_i$ for all $i \in \{1, 2, \dots, d\}$.

Let $j \in \{d+1, \dots, k\}$. First, the domain for $X$ with respect to the unit vector $v_j$ is determined (namely $X$ is contained in a slab $S_j$ of height at most $d\epsilon$ in $v_j$ direction). Secondly, we prove a necessary condition for any translation $x$, such that $a_j + x$ is $\epsilon$-centered between $b_j$ and $b'_j$ (namely $x \in T_j$, a slab of height $\epsilon$ centered at a specific point in the $v_j$-direction). Combining both conditions, the probability that both slabs intersect to finalize the proof.

Let $b^*_l = \frac{b_\ell + b'_\ell}{2}$ be the midpoint between $b_\ell$ and $b'_\ell$. Note that $a_\ell + x$ is $\epsilon$-centered between $b_\ell$ and $b'_\ell$ if $(a_\ell + x) \cdot v_\ell \in \left[b^*_l \cdot v_\ell - \frac{\epsilon}{2}, b^*_l \cdot v_\ell + \frac{\epsilon}{2}\right]$ which is equivalent to

$$(x + a_\ell - b^*_l) \cdot v_\ell \in \left[-\frac{\epsilon}{2}, \frac{\epsilon}{2}\right]. \tag{7}$$

Let $x, x' \in X$, so both $a_i + x$ and $a_i + x'$ are $\epsilon$-centered between $b_i$ and $b'_i$. Then, by Equation (7) and the triangle inequality:

$$|(x - x') \cdot v_i| \leq |(x + a_i - b_i^*)v_i| + |(x' + a_i - b_i^*)v_i| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{8}$$

Now, let's shift our attention to $a_j$, one of the last $k - d$ points. By Lemma 15, it holds that $v_j = \sum_{i=1}^{d} c_{i,j} \cdot v_i$ with $|c_{i,j}| \leq 1$. Combining this with Equation (8) and the triangle inequality,

$$|(x-x')\cdot v_j| = |(x-x')\cdot\left(\sum_{i=1}^{d} c_{i,j} \cdot v_i\right)| \leq \sum_{i=1}^{d}|(x-x')v_i||c_{i,j}| \leq \sum_{i=1}^{d}|(x-x')v_i| \leq \sum_{i=1}^{d} \epsilon = d\epsilon.$$

This implies that $X$ is contained in a slab $S_j$ of height at most $d\epsilon$ in $v_j$ direction, centered somewhere. The position of $S_j$ is independent of $a_{d+1}, \ldots, a_k$, as the condition for $x$ to be in $X$ only depends on the first $d$ points and $v_j$, which consists of $b_j$ and $b'_j$.

Additionally, any translation $x \in \mathbb{R}^d$ satisfies that $a_j + x$ is $\epsilon$-centered between $b_j$ and $b'_j$, if and only if $(x + a_j - b_j^*) \cdot v_j \in \left[-\frac{\epsilon}{2}, \frac{\epsilon}{2}\right]$ by Equation (7). This is equivalent with $x$ belonging to a slab $T_j$ of height $\epsilon$ centered at $b_j^* - a_j$.

By combining the previous results, any translation $x \in \mathbb{R}^d$ satisfying that $a_i + x$ is $\epsilon$-centered between $b_i$ and $b'_i$ for all $i \in \{1, \ldots, d\}$ and $a_j + x$ is $\epsilon$-centered between $b_j$ and $b'_j$ must satisfy two results: first, $x \in X$ and therefore $x \in S_j$, and secondly, $x \in T_j$. Consequently, $x \in S_j \cap T_j$. As $j$ was fixed arbitrarily, it follows that a translation $x \in \mathbb{R}^d$ satisfying that $a_\ell + x$ is $\epsilon$-centered between $b_\ell$ and $b'_\ell$ for all $\ell \in \{1, \ldots, d, \ldots, k\}$ must satisfy that $x \in S_j \cap T_j$ for all $j \in \{d+1, \ldots, k\}$. Notice that a necessary condition for this event is that $S_j \cap T_j \neq \varnothing$ for all $j$.

With these results, we can establish the probability $P$ that the subsets $\{a_1, a_2, \ldots, a_k\}$, and $\{b_1, b_2, \ldots b_k, b'_1, b'_2, \ldots, b'_k\}$ are $(k, \epsilon)$-centerable:

$$\begin{aligned}
P &= P\left(\left(\{a_1, a_2, \ldots, a_k\}, \{b_1, b_2, \ldots b_k, b'_1, b'_2, \ldots, b'_k\}\right) \text{is } (k, \epsilon)\text{-centerable}\right) \\
&= P(\exists x \in \mathbb{R}^d : a_\ell + x \text{ is } \epsilon\text{-centered between } b_\ell \text{ and } b'_\ell \text{ for all } \ell \in \{1, 2, \ldots, k\}) \\
&\leq P(\forall j \in \{d+1, d+2, \ldots, k\} : S_j \cap T_j \neq \varnothing) \\
&= \prod_{j=d+1}^{k} P(S_j \cap T_j \neq \varnothing).
\end{aligned}$$

The last line follows because $S_j \cap T_j$ only uses the randomness of the $a_j$'s, $j \in \{d+1, \ldots, k\}$, which are $k - d$ independent events.

Recall that $S_j$ and $T_j$ are parallel slabs as their respective width is defined with respect to $v_j$. As $S_j$'s position is fixed independently from $a_j$, $S_j \cap T_j \neq \varnothing$ depends on the position of $T_j$ which is centered at $b_j^* - a_j$. As shown in Figure 7, the slab $T_j$ can be translated by a vector $(d+1)\epsilon v_j$ to ensure an intersection with $S_j$. Since, the position of $T_j$ uses the randomness of its center $b_j^* - a_j$, the product $a_j \cdot v_j$ determines the position of the slab in the $v_j$ direction. Thus, the probability that $S_j \cap T_j \neq \varnothing$ is equivalent with the probability that $a_j \cdot v_j$ is in an interval of length $(d+1)\epsilon$. Consequently,

$$P(S_j \cap T_j \neq \varnothing) = P(a_j \cdot v_j \text{ lies in an interval of length } (d+1)\epsilon) \leq (d+1)\epsilon\sqrt{d}\phi,$$

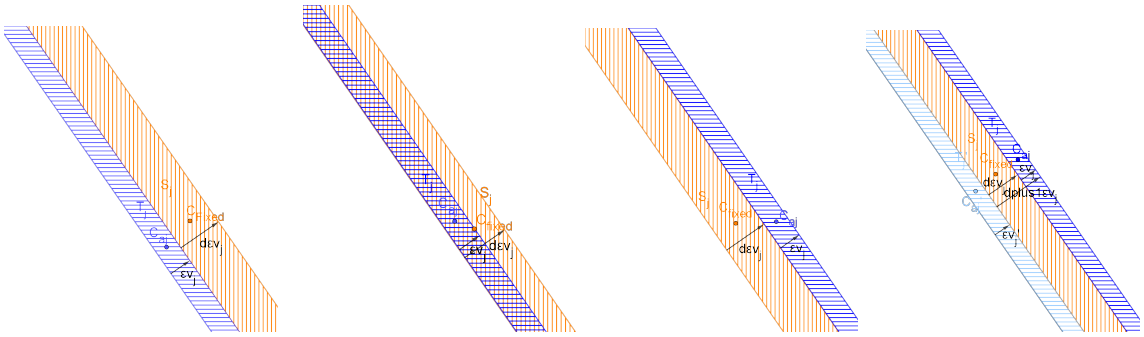
by Lemma 16. Thus,

$$P \leq \left((d+1)\epsilon\sqrt{d}\phi\right)^{k-d}$$

Using the union bound (Lemma B.1), and the fact that there are $n^k m^k m^k$ choices for a combination of $a_1, a_2, \ldots, a_k$, $b_1, b_2, \ldots b_k$ and $b'_1, b'_2, \ldots b'_k$, the final proof follows.

$$
\begin{aligned}
& P((A,B) \text{ is } \epsilon\text{-centerable}) \\
=& P(\exists \text{ distinct } a_1, a_2, \ldots, a_k \in A, b_1, b_2, \ldots, b_k, b'_1, b'_2, \ldots, b'_k \in B, x \in \mathbb{R}^d \text{ such that} \\
& a_i + x \text{ is } \epsilon\text{-centered between } b_i \text{ and } b'_i, \forall i \in \{1, 2, \ldots, k\}) \\
\leq & \sum_{i=1}^{n^k m^{2k}} P \leq n^k m^{2k} \left((d+1)\epsilon\sqrt{d}\phi\right)^{k-d}.
\end{aligned}
$$

□




(A) Begin of the intersection between $S_j$ and $T_J$.

(B) Translate $T_j$ by $\epsilon v_j$. There is still intersection.

(C) Translate $T_j$ by $d\epsilon v_j$. Last line of $S_j$ intersecting $T_j$

(D) Translating $S_j$ by $(d+1)\epsilon v_j$ keeps $S_j$ and $T_j$ intersecting.

FIGURE 7: $S_j$ is a slab fixed somewhere, but the position of $T_j$ depends on the randomness of $a_j$. $S_j$ can be translated by $(d+1)\epsilon$ for $S_j$ and $T_j$ to still intersect (Figures 7a, 7b, 7c and 7d). Thus, the slab centered at $b^*_j - a_j$ must be defined in an interval $(d+1)\epsilon$ in the $v_j$ direction. This is equivalent with the center being defined in a certain interval in the $v_j$ direction. As the center depends on $a_j$'s randomness, $S_j \cap T_j \neq \varnothing$ if $a_j$ is in an interval of length $(d+1)\epsilon$.

The event that $(A, B)$ is $(k, \epsilon)$-centerable (Proposition 17) is called 'Failure 2' and denoted by $F_2$. This completes the analysis of the decrease in potential during an iteration when at least $k$ elements are reassigned: with probability at least $1 - P(F_2)$, the potential decreases by at least $\frac{\delta^2}{n}$.

### 3.3 Proof for the generalized smoothed complexity for the ICP algorithm (Theorem 3)

Combining all the results from the previous subsections, we can prove Theorem 3 using the approach stated at the beginning of this section. The goal is to prove a smoothed upper bound depending on $n, m, d$ and $\phi$. In previous results we declared additional numbers $k, \delta, \epsilon$, which we define now:

*Proof of Theorem 3.* Define the numbers $p$, $k$, $\delta$ and $\epsilon$ as follows:

$$p = \frac{1}{W} = \frac{1}{c(nmd)^d} \qquad \textit{(Theorem 2)},$$
$$k = 2d,$$
$$\delta = \frac{1}{2\phi^{\frac{1}{d}}m^4}p^{\frac{1}{d}}, \text{ and}$$
$$\epsilon = \frac{1}{(d+1)n^2m^4\sqrt{d}\phi}p^{\frac{1}{d}}.$$

Following the strategy described in the beginning of Section 3:

1. Proposition 6 states that the potential $\Phi$ is upper bounded by $\Phi_0 = 4nd$.
2. There are two failure events:
   - 'Failure 1', when $B$ is not $(k,\delta)$-sparse. By Proposition 11,
   $$P(F_1) \leq m^{2k}\phi(2\delta)^d = m^{4d}\phi\left(2\frac{1}{2\phi^{\frac{1}{d}}m^4}p^{\frac{1}{d}}\right)^d = \frac{m^{4d}\phi 2^d p}{2^d\phi m^{4d}} = p.$$
   - 'Failure 2', when $(A,B)$ is $(k,\epsilon)$-centerable. By Proposition 17,
   $$P(F_2) \leq (nm^2)^k((d+1)\epsilon\sqrt{d}\phi)^{k-d}$$
   $$= n^{2d}m^{4d}(d+1)^d(\sqrt{d})^d\phi^d\left(\frac{1}{(d+1)n^2m^4\sqrt{d}\phi}p^{\frac{1}{d}}\right)^d = p.$$

   This means that the failure event $F$ can be written as 'Failure 1 or Failure 2'. Furthermore, its probability can be upper bounded by the union bound (Lemma B.1),
   $$P(F) \leq P(\text{Failure 1}) + P(\text{Failure 2}) \leq 2p.$$

   When there is no failure event, we have that $B$ is $(k,\delta)$-sparse and $(A,B)$ is not $(k,\epsilon)$-centerable. Then the the potential decreases per iteration as follows:

   If there are at most $k$ reassignments (Proposition 9), then
   $$\Delta \geq \frac{\delta^2}{n} = \frac{1}{n}\left(\frac{1}{2\phi^{\frac{1}{d}}m^4}p^{\frac{1}{d}}\right)^2 = \frac{1}{n4\phi^{\frac{2}{d}}m^8}\left(\frac{1}{c(nmd)^d}\right)^{\frac{2}{d}} = \frac{1}{n4\phi^{\frac{2}{d}}m^8c^{\frac{2}{d}}n^2m^2d^2}$$
   $$= \frac{1}{4c^{\frac{2}{d}}n^3m^{10}\phi^{\frac{2}{d}}d^2} \geq \frac{1}{4c^2d^2n^3m^{10}\phi^2} = \frac{1}{q_1(n,m,d,\phi)}.$$

   If there are at least $k$ reassignments (Proposition 14), then
   $$\Delta \geq n\frac{\epsilon^2}{4} = \frac{n}{4}\left(\frac{1}{(d+1)n^2m^4\sqrt{d}\phi}p^{\frac{1}{d}}\right)^2 = \frac{n}{4d(d+1)^2n^4m^8\phi^2}\left(\frac{1}{c(nmd)^d}\right)^{\frac{2}{d}}$$
   $$= \frac{1}{4c^{\frac{2}{d}}d^3(d+1)^2n^5m^{10}\phi^2d^2} \geq \frac{1}{4c^2d^3(d+1)^2n^5m^{10}\phi^2} = \frac{1}{q_2(n,m,d,\phi).}$$

   Thus, if there is no failure event, the decrease of the potential per iteration $\Delta$ is at least $\Delta_0 = \min\left(\frac{1}{q_1(n,m,d,\phi)}, \frac{1}{q_2(n,m,d,\phi)}\right) = \frac{1}{q_2(n,m,d,\phi)}$. Moreover, we have that $\frac{1}{\Delta} \leq \frac{1}{\Delta_0} = q_2(n,m,d,\phi)$.

3. Finally we can determine an upper bound for the expected number of iterations $E[T]$:

$$E(T) \leq W \cdot P(F) + \frac{\Phi_0}{\Delta_0} \leq W \cdot 2p + 4nd \cdot q_2 = 2W\frac{1}{W} + q = q + 2,$$

where $q(n, m, d, \phi) = 4^2c^2d^4(d+1)^2n^6m^{10}\phi^2$, a polynomial in $n$, $m$, $d$, and $\phi$. Hence, the expected number of iterations is upper bounded by a polynomial.

□

# 4 Generalized smoothed analysis of the ICP algorithm with unbounded support

As seen in Section 3, we can find a polynomial upper bound when the probability density function of all points in $A$ and $B$ has bounded support. In this section, we consider probability density functions with unbounded support. We prove that the expected number of iterations of the ICP algorithm can be bounded from above by a polynomial if we add the assumption that the probability distribution has exponentially decreasing tails.

We adopt the notations defined at the beginning of Section 3, except for the following changes.

Let $f_i : \mathbb{R}^d \to [0, \phi]$ for all $i = 1, 2, ..., n$ and $g_j : \mathbb{R}^d \to [0, \phi]$ for all $j = 1, 2, ..., m$ be probability density functions. Define the functions $f$ and $g$ by $f = (f_1, f_2, ..., f_n)$ and $g = (g_1, g_2, ..., g_n)$, respectively. Let the elements of $A$ and $B$ be drawn according to these probability density functions. Furthermore, let $\alpha$, $\beta \in \mathbb{R}^+$, with $\beta \neq 0$ and $\alpha \geq \frac{1}{W}$ ($W$ being the maximal number of iterations as defined in Theorem 2). Then, we assume that, for all $z \in A \cup B$, we have $P(\|z\| \geq t) \leq \alpha \exp(-\beta(2t)^{d-1})$, for all $t \in \mathbb{R}^+$.

Additional assumptions or deviant notations are specified in a lemma or proposition if needed.

**Theorem 18** (Generalized smoothed complexity for the ICP algorithm)**.** *Let the elements of $A$ and $B$ be defined as described above. Then, the expected number of iterations of the ICP algorithm is bounded from above by a polynomial in $n$, $m$, $d$, $\frac{1}{\alpha}$, $\beta$ and $\phi$.*

To prove this theorem, some of the lemmas and propositions of Section 3 can be reused. However, Proposition 6, Lemma 16 and Proposition 17 need to be adapted.

**Proposition 19** (Adaptation of Proposition 6)**.** *Let $t \in \mathbb{R}$, and assume that the norm of all points in $A$ and $B$ is less than $t$. Then, we have an upper bound for the potential $\Phi$,*

$$\Phi = \sum_{a \in A} \|a + x - N_B(a)\|^2 \leq 16nt^2.$$

*Proof.* In the ICP algorithm, the translation is obtained as follows, $x = \frac{1}{|A|}\sum_{a \in A}(N_B(a) - a)$. For all $a \in A$, we have $\|a\| < t$ and $N_B(a) \in B$, so $\|N_B(a)\| < t$. Thus,

$$\|x\| = \left\| \frac{1}{|A|} \sum_{a \in A} (N_B(a) - a) \right\| \leq \frac{1}{n} \sum_{a \in A} \|N_B(a) - a\| \leq \frac{1}{n} \sum_{a \in A} (\|N_B(a)\| + \|a\|) < t+t = 2t.$$

Then, we have that

$$\Phi = \sum_{a \in A} \|a + x - N_B(a)\|^2 \leq \sum_{a \in A} (\|a\| + \|x\| + \|N_B(a)\|)^2 = n(4t)^2 = 16nt^2.$$

□

**Lemma 20** (Adaptation of Lemma 16)**.** *Let $t \in \mathbb{R}$ and let $z = (z_1, z_2, \ldots, z_d)$ be a point in $A$ with $\|z\| < t$. Let $v = (v_1, v_2, \ldots, v_d) \in \mathbb{R}^d$ be a unit vector. Let $I \subset \mathbb{R}$ be an interval of length $\epsilon$. Then, the probability that the dot product $z \cdot v$ lies in $I$ is at most $\sqrt{d}\epsilon\phi(2t)^{d-1}$,*

$$P(z \cdot v \in I) \leq \sqrt{d}\epsilon\phi(2t)^{d-1}.$$

*Proof.* Let $v_i$ be the component of $v$ with the largest absolute value. By the proof of Lemma 16, we know that $\frac{1}{|v_i|} \leq \sqrt{d}$ and $P(z \cdot v \in I) = P\left(z_i \in \left[\frac{1}{v_i}c^*, \frac{1}{v_i}e^*\right]\right)$, with $e^* - c^* = \epsilon$. Since $\|z\| < t$, this means that the absolute value of every component of $z$ is less than $t$. Thus, every component lies in the range $[-t, t]$. Recall that $z$ is drawn according to some probability density function $f$ of the form $\mathbb{R}^d \to [0, \phi]$. Then,

$$\begin{aligned}
P(z \cdot v \in I) &\leq \int_{-t}^{t}\int_{-t}^{t}\ldots\int_{-t}^{t}\int_{\frac{1}{v_i}c^*}^{\frac{1}{v_i}e^*} f(z_1 = x_1, z_2 = x_2 \ldots, z_d = x_d)\, dx_i\, dx_1 \ldots\, dx_d \\
&\leq \int_{-t}^{t}\int_{-t}^{t}\ldots\int_{-t}^{t}\frac{1}{v_i}\epsilon\phi\, dx_1 \ldots\, dx_d \\
&= (2t)^{d-1}\frac{1}{v_i}\epsilon\phi \leq \sqrt{d}\epsilon\phi(2t)^{d-1}.
\end{aligned}$$

□

**Proposition 21** (Adaptation of Proposition 17)**.** *Let $t \in \mathbb{R}$, and assume that the norm of all points in $A$ and $B$ is less than $t$. Then, $(A, B)$ is $(k, \epsilon)$-centerable with probability*

$$P(\textit{$(A, B)$ is $(k, \epsilon)$-centerable}) < (nm^2)^k\left((d+1)\epsilon\sqrt{d}\phi(2t)^{d-1}\right)^{k-d}.$$

*Proof.* The proof for this theorem is equivalent to the proof of Proposition 17, except that not the result of Lemma 16 is used but the result of Lemma 20. Then,

$$P(a_j \cdot v_j \text{ lies in an interval of length } (d+1)\epsilon) \leq (d+1)\epsilon\sqrt{d}\phi(2t)^{d-1}.$$

Thus, we have

$$P \leq \left((d+1)\epsilon\sqrt{d}\phi(2t)^{d-1}\right)^{k-d}.$$

Using the union bound (Lemma B.1), and the fact that there are $n^k m^k m^k$ choices for a combination of $a_1, a_2, \ldots, a_k$, $b_1, b_2, \ldots b_k$ and $b'_1, b'_2, \ldots b'_k$, the final proof follows.

$$\begin{aligned}
&P((A, B) \text{ is } \epsilon\text{-centerable}) \\
=&P(\exists \text{ distinct } a_1, a_2, \ldots, a_k \in A, b_1, b_2, \ldots, b_k, b'_1, b'_2, \ldots, b'_k \in B, x \in \mathbb{R}^d \text{ such that} \\
&a_i + x \text{ is } \epsilon\text{-centered between } b_i \text{ and } b'_i, \forall i \in \{1, 2, \ldots, k\}) \\
\leq &\sum_{i=1}^{n^k m^{2k}} P \leq n^k m^{2k}\left((d+1)\epsilon\sqrt{d}\phi(2t)^{d-1}\right)^{k-d}.
\end{aligned}$$

□

By 'Failure 2' ($F_2$), we now denote the event that, given the norm of all points in $A$ and $B$ is less than some $t \in \mathbb{R}$, $(A, B)$ is $(k, \epsilon)$-centerable (Proposition 21). After having revisited the propositions that use the bounded support of the points in $A$ and $B$ in Section 3, we can state the proof of Theorem 18. All propositions used from Section 3 are to be understood with the new definitions of the sets $A$ and $B$.

*Proof of Theorem 18.* Define the numbers $p$, $k$, $\delta$, $\epsilon$ and $t$ as follows:

$$
\begin{aligned}
&p = \frac{1}{W} = \frac{1}{c(nmd)^d} && \textit{(Theorem 2)},\\
&k = 2d,\\
&t = \frac{1}{2}\sqrt[d-1]{\frac{-1}{\beta}\ln\frac{p}{(m+n)\alpha}} = \frac{1}{2}\sqrt[d-1]{\frac{1}{\beta}\left(d\ln nmd\right) + \ln\left(c\alpha(m+n)\right)},\\
&\delta = \frac{1}{2\phi^{\frac{1}{d}}m^4}p^{\frac{1}{d}}, \text{ and}\\
&\epsilon = \frac{1}{(d+1)n^2m^4\sqrt{d}\phi(2t)^{d-1}}p^{\frac{1}{d}}.
\end{aligned}
$$

We denote three failure events:

- A point $z$ of $A$ or $B$ has a norm greater than $t$. This happens with probability

$$
\begin{aligned}
P(\|z\| \geq t) &\leq \alpha\exp\left(-\beta(2t)^{d-1}\right) = \alpha\exp\left(-\beta\left(2\frac{1}{2}\sqrt[d-1]{\frac{-1}{\beta}\ln\frac{p}{(m+n)\alpha}}\right)^{d-1}\right)\\
&= \frac{p}{m+n}.
\end{aligned}
$$

  We denote by 'Failure 0' ($F_0$), if any point of $A$ or $B$ has a norm greater than $t$. By the union bound,

$$
P(F_0) \leq (m+n)\frac{p}{m+n} = p
$$

- 'Failure 1', that the set $B$ is not $(k, \delta)$-sparse. By Proposition 11,

$$
P(F_1) \leq m^{2k}\phi(2\delta)^d = m^{4d}\phi\left(2\frac{1}{2\phi^{\frac{1}{d}}m^4}p^{\frac{1}{d}}\right)^d = \frac{m^{4d}\phi 2^d p}{2^d\phi m^{4d}} = p.
$$

- 'Failure 2', namely given that all points of $A$ and $B$ have a norm less than $t$, then $(A, B)$ is $(k, \epsilon)$-centerable. By Proposition 21,

$$
\begin{aligned}
P(F_2) &\leq (nm^2)^k((d+1)\epsilon\sqrt{d}\phi(2t)^{d-1})^{k-d}\\
&= n^{2d}m^{4d}(d+1)^d(\sqrt{d})^d\phi^d\left((2t)^{d-1}\right)^d\left(\frac{1}{(d+1)n^2m^4\sqrt{d}\phi(2t)^{d-1}}p^{\frac{1}{d}}\right)^d\\
&= p.
\end{aligned}
$$

The probability of the failure event $F$ that none of these failures occur can be upper bounded by $P(F) \leq P(F_0) + P(F_1) + P(F_2) \leq 3p$. So with probability $1 - 3p$, we have that all points in $A$ and $B$ have a norm less than $t$, that $B$ is $(k, \delta)$-sparse and $(A, B)$ is not $(k, \epsilon)$-centerable. Given these conditions, the following decrease of the potential per iteration can be established:

- If there are at most $k$ reassignments (Proposition 9):

$$\Delta \geq \frac{\delta^2}{n} = \frac{1}{n}\left(\frac{1}{2\phi^{\frac{1}{d}}m^4}p^{\frac{1}{d}}\right)^2 \geq \frac{1}{4c^2d^2n^3m^{10}\phi^2} = \frac{1}{q_1(n,m,d,\phi)}.$$

- If there are at least $k$ reassignments, we first define

$$(2t)^{d-1} = \left(2\frac{1}{2}\sqrt[d-1]{\frac{-1}{\beta}\ln\frac{p}{(m+n)\alpha}}\right)^{d-1} = \frac{-1}{\beta}\ln\frac{1}{(m+n)\alpha c(nmd)^d}$$
$$= \frac{1}{\beta}\left(d\ln(nmd) + \ln(c(m+n)\alpha)\right) \leq \frac{1}{\beta}\left(d^2nm + c(m+n)\alpha\right).$$

  Then, we can establish a lower bound on the drop of the potential $\Delta$ (Proposition 14):

$$\Delta \geq n\frac{\epsilon^2}{4} = \frac{n}{4}\left(\frac{1}{(d+1)n^2m^4\sqrt{d}\phi(2t)^{d-1}}p^{\frac{1}{d}}\right)^2 \geq \frac{1}{4c^2d^3(d+1)^2n^5m^{10}\phi^2}\left(\frac{1}{(2t)^{d-1}}\right)^2$$
$$\geq \frac{1}{4c^2d^3(d+1)^2n^5m^{10}\phi^2}\left(\frac{1}{\frac{1}{\beta}\left(d^2nm + c(m+n)\alpha\right)}\right)^2$$
$$= \frac{1}{q_2\left(n,m,d,\phi,\alpha,\frac{1}{\beta}\right)}.$$

Thus, if there is no failure event, the decrease of the potential per iteration $\Delta$ is at least $\Delta_0 = \min\left(\frac{1}{q_1(n,m,d,\phi)}, \frac{1}{q_2\left(n,m,d,\phi,\alpha,\frac{1}{\beta}\right)}\right)$. Define the polynomial

$$q_*\left(n,m,d,\phi,\alpha,\frac{1}{\beta}\right) = 4c^2d^3(d+1)^2n^5m^{10}\phi^2\left(\frac{1}{\beta^2}\left(d^2nm + c(m+n)\alpha\right)^2 + 1\right) \geq \max(q_1, q_2).$$

Then, we have that

$$\frac{1}{\Delta} \leq \frac{1}{\Delta_0} = \max(q_1, q_2) \leq q_*\left(n,m,d,\phi,\alpha,\frac{1}{\beta}\right).$$

Furthermore, if the norm of all points in $A$ and $B$ is bounded by $t$, then the initial potential $\Phi$ can be upper bounded by $\Phi_0 = 16nt^2$ (Proposition 19).

Finally we can determine an upper bound for the expected number of iterations $E[T]$:

$$\begin{aligned} E(T) &\leq W \cdot P(F) + \frac{\Phi_0}{\Delta_0} \leq W \cdot 3p + 16nt^2 q_* \\ &= 3W\frac{1}{W} + 16n\frac{1}{4}\left(\frac{1}{\beta}\left(d\ln\left(nmd\right) + \ln\left(c(m+n)\alpha\right)\right)\right)^{\frac{2}{d-1}} q_* \\ &\leq 3 + 4n\frac{1}{\beta^2}\left(d^2nmd + c(m+n)\alpha\right)^2 q_* \\ &= q\left(n, m, d, \phi, \alpha, \frac{1}{\beta}\right) + 3, \end{aligned}$$

where $q\left(n, m, d, \phi, \alpha, \frac{1}{\beta}\right) = 4n\frac{1}{\beta^2}\left(d^2nmd + c\alpha\right)^2 q_*$ is a polynomial in $n$, $m$, $d$, $\phi$, $\alpha$ and $\frac{1}{\beta}$. □

Hence, the expected number of iterations is upper bounded by a polynomial.

# 5 Discussion and Recommendations

To generalize the smoothed analysis by Arthur and Vassilvitskii, two different perturbation models are introduced. In Section 3, the analysis is generalized by drawing the points of the source and reference set not from a Gaussian distribution but according to arbitrary density functions bounded by the perturbation parameter. However, to facilitate calculations, the restriction was added that all elements are situated inside the unit hypercube. To remove this limitation, in Section 4, all elements are drawn from arbitrary probability distributions with unbounded support. However, to achieve a polynomial upper bound a new restriction on the tails of the distribution was required.

A future goal would be to make further generalisations on the smoothed analysis, especially when the probability density functions have unbounded support. In Section 4, the bottleneck of the analysis is Lemma 20 which introduces a term having the dimension as exponent. Consequently, the condition on the tails of the probability distribution must be rather strict. Furthermore, Lemma 20 only makes use of the upper bound on the probability density function so far and not also of the exponential tails in the distribution. For this reason, we believe that the result could potentially be improved by replacing the condition on the norm of the points in $A$ and $B$ by a condition on the probability density function.

Both smoothed upper bounds on the expected number of iterations, as well as the bound established by Arthur and Vassilvitskii are polynomial. This supports the observations that in practice the ICP algorithm is efficient and well-used. However, note that even though the bounds on the expected number of iterations are polynomial, their degrees are very large. Even though polynomial bounds are a first step to reconcile theory and practice, we expect that this gap could still be narrowed. To achieve such a substantial improvement, either new techniques need to be found to acquire better results on this type of proof structure, or an entirely new proof concept would have to be developed.

When studying the lower bound on the worst-case running time of the ICP algorithm, established by Arthur and Vassilvitskii [1], an unsuccessful attempt was made to get experimental results on the number of iterations of their constructed worst-case instance. The implementation of a worst-case would have allowed us to examine the practical significance of small perturbations in the data set (i.e. how significant is the decrease in the

number of iterations of the ICP algorithm). Nonetheless, we think that such experiments could lead to useful insights into the functioning of the ICP algorithm and strengthen the reasoning to use smoothed analysis.

# 6 Conclusion

After the worst-case upper bound of the ICP algorithm's running time determined by Ezra et al. [8], and the smoothed analysis by Arthur and Vassilvitskii [1], this paper provides additional results on the theoretical analysis of this local search algorithm. In Section 3, a proof is given for a polynomial upper bound when the elements of the source set $A$ and the reference set $B$ are drawn according to probability density functions of the form $[0,1]^d \to [0,\phi]$. To investigate the behaviour of the algorithm when the probability density functions have unbounded support, Section 4 shows the generalized smoothed analysis when the probability density functions are of the form $\mathbb{R}^d \to [0,\phi]$. To achieve a polynomial upper bound on the expected number of iterations, it is required that the probability distribution has exponentially decreasing tails.

As the ICP algorithm has a polynomial smoothed complexity, it is expected to work well in practice. This means that theoretical advances are one step closer to explain the observations about the algorithm's efficiency. Nonetheless, more research into different perturbation models and into reducing the polynomial's large degrees could lead to an even more profound understanding of the ICP algorithm.

All in all, the obtained results of the smoothed analysis support the observations that the ICP algorithm performs well in practice and help theoretical advances to get a step closer of backing this up.

# References


[1] David Arthur and Sergei. Vassilvitskii. Worst-Case and Smoothed Analysis of the ICP Algorithm, with an Application to the k-Means Method. *SIAM Journal on Computing*, 39(2):766–782, 2009.

[2] R Beier and B Vöcking. Typical Properties of Winners and Losers in Discrete Optimization. *SIAM Journal on Computing*, 35(4):855–881, 2006.

[3] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb 1992.

[4] Kenneth George Binmore and Kenneth George Binmore. *The Foundations of Topological Analysis: A Straightforward Introduction: Book 2 Topological Ideas*. CUP Archive, 1980.

[5] Tobias Brunsch, Kamiel Cornelissen, Bodo Manthey, and Heiko Röglin. Smoothed Analysis of Belief Propagation for Minimum-Cost Flow and Matching. *Journal of Graph Algorithms and Applications*, 17(6):647–670, 2013.

[6] Radu Curticapean and Marvin Künnemann. A quantization framework for smoothed analysis of euclidean optimization problems. *Algorithmica*, 73(3):483–510, 2015.

[7] Shaoyi Du, Nanning Zheng, Lei Xiong, Shihui Ying, and Jianru Xue. Scaling iterative closest point algorithm for registration of m–d point sets. *Journal of Visual Communication and Image Representation*, 21(5-6):442–452, 2010.

[8] Esther Ezra, Micha Sharir, and Alon Efrat. On the performance of the ICP algorithm. *Computational Geometry*, 41(1-2):77–93, 2008.

[9] Shao-Ya Guan, Tian-Miao Wang, Cai Meng, and Jun-Chen Wang. A review of point feature based medical image registration. *Chinese Journal of Mechanical Engineering*, 31(1):76, Aug 2018.

[10] Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. A local search approximation algorithm for k-means clustering. *Computational Geometry*, 28(2-3):89–112, 2004.

[11] Higinio Mora, Jerónimo M Mora-Pascual, Alberto Garcia-Garcia, and Pablo Martinez-Gonzalez. Computational analysis of distance operators for the iterative closest point algorithm. *PloS one*, 11(10):e0164694, 2016.

[12] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the ICP algorithm. In *3dim*, volume 1, pages 145–152, 2001.

[13] Daniel A. Spielman and Shang-Hua Teng. Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time. *Journal of the ACM*, 51(3):385–463, 2004.

[14] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis: an attempt to explain the behavior of algorithms in practice. *Commun. ACM*, 52(10):76–84, 2009.

[15] Eric W. Weisstein. Half-space. From MathWorld—A Wolfram Web Resource. Last visited on 22/1/2020.

[16] Eric W. Weisstein. Multiset. From MathWorld—A Wolfram Web Resource. Last visited on 22/1/2020.

# A Notations, definitions and additional lemmas

## A.1 Table of symbols

TABLE 1: Table of symbols

| Symbol | Notes |
|---|---|
| $d \in \mathbb{N}$ | Dimension |
| $A \subseteq \mathbb{R}^d$ | Source set of the ICP algorithm: points from $A$ should be matched to another pattern minimizing the potential. |
| $B \subseteq \mathbb{R}^d$ | Reference set of the ICP algorithm: Points in $A$ need to be matched to the set $B$ minimizing the potential. Multiple elements of $A$ can be assigned to the same element in $B$. |
| $n$ | Number of elements in $A$, $n = \|A\|$ |
| $m$ | Number of elements in $B$, $m = \|B\|$ |
| $W$ | Maximum number of iterations: The ICP algorithm terminates in at most $c(\|A\|, \|B\|, d)^d$ iterations (Theorem 2). |
| $x \in \mathbb{R}^d$ | Translation (ICP algorithm), $x = \frac{1}{\|A\|} \sum_{a \in A} (N_B(a) - a)$ in step 6 (Algorithm 1). |
| $N_B$ | Nearest neighbour function: $N_B : A \to B$ minimizing the potential during the ICP algorithm (Steps 3-5 in Algorithm 1). |
| $\Phi$ | Potential of the ICP algorithm, $\Phi = \sum_{a \in A} \|a + x - N_B(a)\|$ |
| $c(S)$ | Center of mass of a set $S$ , $c(S) = \frac{1}{\|S\|} \sum_{s \in S} s$ |
| $\phi$ | Perturbation parameter: upper bound of the generalized probability density function. functions used in Theorem 3. |
| $\Phi_0$ | Upper bound on the potential with high probability, $\Phi_0 = 4\|A\|d$. on $A$, $B$, $d$ and $\phi$. |
| $\Delta$ | Decrease of the potential $\Phi$ per iteration. |
| $\Delta_0 > 0$ | Lower bound of $\Delta$ with high probability, |
| $k \in \mathbb{N}$ | In the proof of Theorem 3, we first analyse an iteration when there are at most $k$ changes in $N_B$ and when there are at least $k$ such changes. |
| $\delta \in \mathbb{R}$ | Variable introduced by the proof of Theorem 3 to define the distance measure in Definition 8. |
| $t \in \mathbb{R}^+$ | In Section 4, the norm of all points in $A$ and $B$ is assumed to be less than $t$, otherwise we have a failure. |
| $\alpha \geq p$ | Perturbation parameter in Section 4. |
| $\beta \in \mathbb{R}^+$ | Perturbation parameter in Section 4. |
| $\epsilon \in \mathbb{R}$ | Variable introduced by the proof of Theorem 3 to define the distance measure in Definition 12. |
| $F_1$ | The event 'Failure 1' that $B$ is not $(k, \delta)$-sparse. |
| $F_2$ | The event 'Failure 2' in Section 3: $(A, B)$ is $(k, \epsilon)$-centerable. In Section 4: given that the norm of all points in $A$ and $B$ is less than some $t \in \mathbb{R}$, $(A, B)$ is $(k, \epsilon)$-centerable. |
| $p$ | Failure probability |
| $T$ | Number of iterations of the ICP algorithm. |
| $f$ | $f = (f_1, f_2, \ldots, f_n)$ with $f_i$ a probability density function for all $i = 1, 2, \ldots, n$. The points in $A$ are drawn according to these functions. |
| $g$ | $g = (g_1, g_2, \ldots, g_m)$ with $g_j$, a probability density function for all $j = 1, 2, ..., m$. The points in $B$ are drawn according to these functions. |
| $\|v\|$ | Euclidean norm of a vector $v \in \mathbb{R}^d$ with $d \geq 2$. If $v$ is a d-dimensional point of $A$ and $B$, $\|v\|$ denotes the distance between the origin and the point. |
| $\|v - v'\|$ | Euclidean distance between two points $v, v' \in \mathbb{R}$. |

### A.2 Definitions and notations

**Definition A.1** (Center of mass)**.** The center of mass of a set $S$ is defined as $c(S) = \frac{1}{|S|}\sum_{s\in S} s$.

**Definition A.2** (Iteration)**.** In this paper, an iteration is defined to be the execution of steps 3-6 in algorithm 1.

**Definition A.3** (Multiset)**.** "A set-like object in which order is ignored, but multiplicity is explicitly significant. Therefore, multisets $\{1,2,3\}$ and $\{2,1,3\}$ are equivalent, but $\{1,1,2,3\}$ and $\{1,2,3\}$ differ" [16].

**Definition A.4** (Hyperplane bisecting two points in $\mathbb{R}^d$)**.** In $\mathbb{R}^d$ , a hyperplane is a subspace whose dimension equals $d-1$. For example, in a 3-dimensional space, hyperplanes are the 2-dimensional planes. The hyperplane bisecting two points $a, b \in \mathbb{R}^d$ is the hyperplane that contains the middle point $\frac{a+b}{2}$ and is orthogonal to to the vector $b-a$.

**Definition A.5** (Half-Space)**.** "A half-space is that portion of an $d$-dimensional space obtained by removing that part lying on one side of an $(d-1)$-dimensional hyperplane" [15].

## B Lemmas and Theorems

**Lemma B.1.** *Let $S \subseteq \mathbb{R}^d$ be finite and let $x \in \mathbb{R}^d$. Define $c(S)$ to be the center of mass of $S$. Then,*

$$\sum_{s\in S} \|x-s\|^2 = \sum_{s\in S} \|s-c(S)\|^2 + |S|\,\|c(S)-x\|^2 .$$

*Proof.* Imitating the proof of Kanungo et al. [10, Lemma 2.1] and using that an element's squared norm is equal to the inner product with itself, we obtain

$$\begin{aligned}
\sum_{s\in S} \|x-s\|^2 &= \sum_{s\in S} (s-x)(s-x) \\
&= \sum_{s\in S} \Big((s-c(S)) + (c(S)-x)\Big)\Big((s-c(S)) + (c(S)-x)\Big) \\
&= \sum_{s\in S} \Big((s-c(S))(s-c(S)) + 2(s-c(S))(c(S)-x) + (c(S)-x)(c(S)-x)\Big) \\
&= \sum_{s\in S} \|s-c(S)\|^2 + \sum_{s\in S} 2(s-c(S))(c(S)-x) + \sum_{s\in S} \|c(S)-x\|^2 \\
&= \sum_{s\in S} \|s-c(S)\|^2 + 2(c(S)-x)\sum_{s\in S}(s-c(S)) + |S|\,\|c(S)-x\|^2 .
\end{aligned}$$

Since $c(S) = \frac{1}{|S|}\sum_{s\in S} s$, we have that $\sum_{s\in S}(s-c(S)) = \left(\sum_{s\in S} s\right) - |S|c(S) = \sum_{s\in S} s - |S|\frac{1}{|S|}\sum_{s\in S} s = 0$, so the final equation becomes

$$\sum_{s\in S} \|x-s\|^2 = \sum_{s\in S} \|s-c(S)\|^2 + |S|\,\|c(S)-x\|^2 .$$

□

**Lemma B.2.** *Let $A, B \subseteq \mathbb{R}^d$. Let $N_B : A \to B$ be the nearest neighbour function for the ICP algorithm. For a fixed $N_B$, the potential $\Phi$ is minimized by choosing the translation*

$$x = \frac{1}{|A|} \sum_{a \in A} (N_B(a) - a).$$

*Proof.* Let $S$ be the multiset $\{N(a) - a \mid a \in A\}$ and $c(S)$ be the center of mass of $S$. Then, we can write

$$\sum_{a \in A} \|a + x - N(a)\|^2 = \sum_{s \in S} \|x - s\|^2 = \sum_{s \in S} \|s - c(S)\|^2 + |S| \, \|c(S) - x\|^2 ,$$

by Lemma B.1. Then, the first term $\sum_{s \in S} \|s - c(S)\|^2$ is independent of $x$. Thus, $\Phi$ is minimized if and only if $|S| \cdot \|c(S) - x\|^2$ is minimized. Hence, the translation $x$ to minimize $\Phi$ is equal to $c(S) = \frac{1}{|A|} \sum_{a \in A} (N(a) - a)$,

$$x = c(S) = \frac{1}{|S|} \sum_{s \in S} s = \frac{1}{|A|} \sum_{a \in A} (N(a) - a).$$

□

**Lemma B.3.** *Let $v \in \mathbb{R}^d$ be a unit vector. Let $v_i$ be the component of $v$ with the largest absolute value, $v_i = \text{argmax}_{v \in V} |v|$. Then $|v_i| \geq \frac{1}{\sqrt{d}}$.*

*Proof.* We have

$$\begin{aligned} & 1 = \|v\| = \sqrt{v_1^2 + v_2^2 \cdots + v_d^2} \leq \sqrt{v_i^2 + v_i^2 + \cdots + v_i^2} = \sqrt{d v_i^2} \\ \iff & 1^2 \leq \sqrt{d} |v_i| \\ \iff & |v_i| \geq \frac{1}{\sqrt{d}} \end{aligned}$$

□

Here a collection of commonly known results that are used in this paper:

**Theorem B.1** (Union bound, Boole's inequality)**.** *Let $E_1$, ...., $E_n$ be any events. Then*

$$P(E_1 \vee E_2 \vee ... \vee E_n) = P(\exists i \in 1, ..., n : E_i) \leq \sum_{i=1}^{n} P(E_i)$$

**Lemma B.4.** *For all $x > 0$, we know that $\ln x < x$.*