

Supplementary material for the article: Comparison-Based Random Forests

Siavash Haghiri, Damien Garreau, Ulrike von Luxburg

In this supplementary material, we provide a complete proof of Theorem 2 and additional experiments. The main arguments of this proof are collected in Section 1, while all auxiliary results can be found in Section 2. Section 3 contains the supplemental experiments.

1 Main proofs

Let us recall our main assumptions on the data and the central result regarding the consistency of the continuous comparison tree.

Assumption 1.1 (Bounded density on the unit cube). The random variable $X \in [0, 1]^d$ has density f with respect to the Lebesgue measure on $[0, 1]^d$. Additionally, there exist constants $0 < f_{\min} \leq f_{\max} < +\infty$ such that

$$\forall x \in \mathcal{X}, \quad f_{\min} \leq f(x) \leq f_{\max}.$$

Theorem 1.1 (Consistency of comparison trees). *Assume that X satisfies Assumption 1.1. Then the partitioning classification rule associated to $T^0(\alpha \log n)$ is consistent for any $0 < \alpha < 1/\log 2$.*

As explained in the paper, the proof of Theorem 1.1 follows from Devroye et al. (1996, Theorem 6.1). It first requires control on the number of samples in the leaves of the tree. This is resolved in Section 1.2. Second, we need to bound the diameter of the leaves of the tree, which is achieved in Section 1.3. Before turning to these proofs, we state and prove the key result used to control the diameter (Proposition 1.1) in the next section.

Let us now introduce some additional notation. For any convex compact subset A , we define π_A as the orthogonal projection on A . For any given $x \in \mathbb{R}^d$ and $r > 0$, we define $\mathcal{B}(x, r)$ as the closed ball of center x and radius r . Namely,

$$\mathcal{B}(x, r) := \{y \in \mathbb{R}^d \mid \|x - y\| \leq r\}.$$

The sphere of center x and radius r is the boundary of $\mathcal{B}(x, r)$, that is,

$$\mathcal{S}(x, r) := \{y \in \mathbb{R}^d \mid \|x - y\| = r\}.$$

When working in dimension 2, $\mathcal{S}(x, r)$ is simply the circle of center x and radius r , denoted by $\mathcal{C}(x, r)$. We call *annulus* the (closed) set of points comprised between two concentric spheres, that is, for any $x \in \mathbb{R}^d$, $r_1, r_2 > 0$,

$$\mathcal{A}(x, r_1, r_2) := \{y \in \mathbb{R}^d \mid r_1 \leq \|x - y\| \leq r_2\}.$$

If $r_1 \leq 0$, we set $\mathcal{A}(x, r_1, r_2) = \mathcal{B}(x, r_2)$.

1.1 Control of the cell diameter

Let us recall Proposition 1.1, the key result needed for proving that the diameter of the comparison-tree leaves goes to zero in probability.

Proposition 1.1 (Diameter control). *Assume that Assumption 1.1 holds. Let C be a cell of $T^0(X)$ such that $\text{diam}(C) \leq D$. Then the probability that there exists a descendant of C which is more than k levels below and yet has diameter greater than $D/2$ is at most $N_{f,d}(N_{f,d} + 1)\gamma_{f,d}^k/2$, where $0 < N_{f,d}$ and $0 < \gamma_{f,d} < 1$ are constants depending only on d , f_{\min} , and f_{\max} .*

Proposition 1.1 is an analogous of Lemma 12 in Dasgupta and Freund (2008). In plain words, it states that for any cell of the continuous comparison tree, the diameter of any descendant at least k levels below is halved with high probability depending on k . Our proof follows closely that of Dasgupta and Freund (2008, Lemma 12), the main difference being in the auxiliary lemmas used to control the probability of certain events, due to the radically different nature of the random tree that we consider.

Proof. Consider a cover of C by balls of radius $r = D/c_r$, with $c_r := 2^6 \cdot d \cdot 25^d \cdot \frac{f_{\max}^2}{f_{\min}^2}$. According to Shalev-Shwartz and Ben-David (2014, Section 27.1), at most

$$\left(\frac{2D\sqrt{d}}{r}\right)^d = \left(2^7 \cdot d^{3/2} \cdot 25^d \cdot \frac{f_{\max}^2}{f_{\min}^2}\right)^d =: N_{d,f}$$

such balls are needed, since $\text{diam}(C) \leq D$.

Fix any pair of balls B, B' from this cover whose centers are at distance at least $D/2 - r$ from one another. Given any x and y , we say that the split according to $\Delta(x, y)$ is a *good cut* if it cleanly separates B from B' , i.e., if $B \subset H_x$ and $B' \subset H_y$ or $B' \subset H_x$ and $B \subset H_y$. If the split cuts both B and B' , that is, $B \cap \Delta(x, y) \neq \emptyset$ and $B' \cap \Delta(x, y) \neq \emptyset$, we say that it is a *bad cut*. See Figure 1 for illustration.

For any $k \geq 1$, let p_k be the probability that there is some cell k levels below C which contains points from both B and B' . We write

$$\begin{aligned} p_k &\leq \mathbb{P}(\text{top split is a good cut}) \cdot 0 + \mathbb{P}(\text{top split is a bad cut}) \cdot 2p_{k-1} \\ &\quad + \mathbb{P}(\text{all other split configurations}) \cdot p_{k-1} \\ &\leq (1 + \mathbb{P}(\text{top split is a bad cut}) - \mathbb{P}(\text{top split is a good cut})) p_{k-1}. \end{aligned}$$

Since $d \geq 1$ and $c_r > 50$, according to Lemma 2.1 and 2.2,

$$\begin{aligned} \mathbb{P}(\text{top split is a bad cut}) - \mathbb{P}(\text{top split is a good cut}) &\leq \frac{f_{\max}}{f_{\min}} \cdot \frac{64d}{c_r} - 2 \cdot \frac{f_{\min}}{f_{\max}} \cdot \frac{1}{25^d} \\ &= -\frac{f_{\min}}{f_{\max}} \cdot \frac{1}{25^d} < 0. \end{aligned}$$

Set $\gamma_{f,d} := 1 - \frac{f_{\min}}{f_{\max}} \cdot \frac{1}{25^d}$, we just showed that

$$p_k \leq \gamma_{f,d} p_{k-1}. \tag{1.1}$$

Since $p_0 = 1$, we deduce that $p_k \leq \gamma_{f,d}^k$. We conclude by a union bound over all the pairs from the cover that are at the prescribed minimum distance from each other. \square

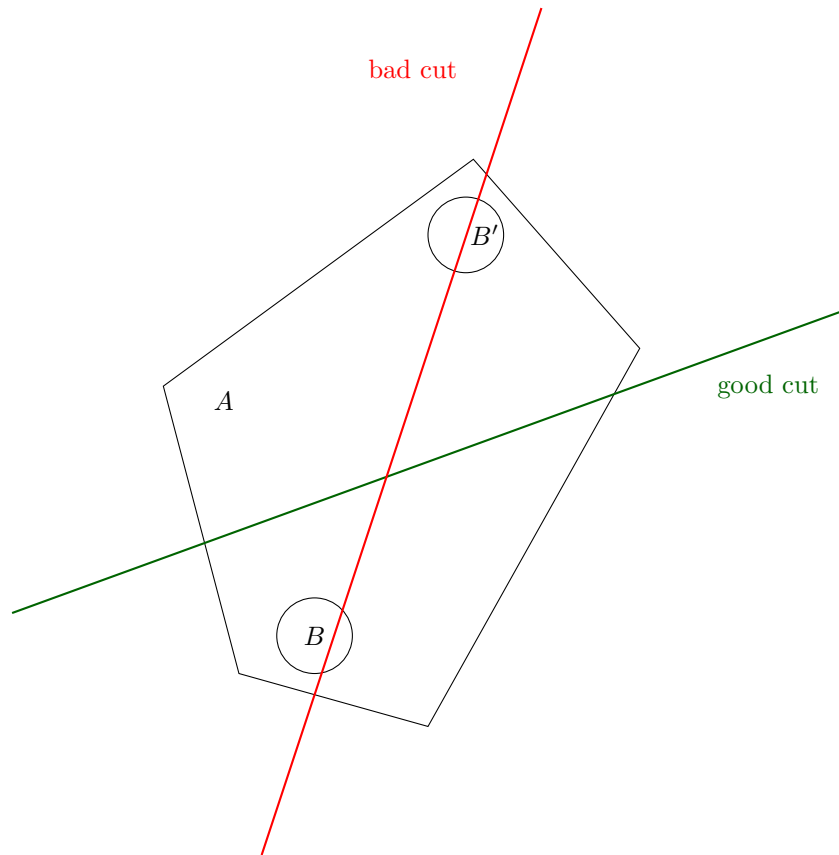


Figure 1: Good cuts and bad cuts. The current cell A contains B and B' , two faraway balls of small radius—with respect to the diameter of A . A *good cut* (in green) cleanly separates B and B' , whereas a *bad cut* (in red) intersects both.

Note that the main difference with the proof of Lemma 12 in Dasgupta and Freund (2008) comes from Eq. (1.1). Namely, in the setting of Dasgupta and Freund (2008), $\gamma_{f,d}$ is a constant that does not depend on the dimension. The dependency on the dimension in our case is due to the lower bound on the probability of a good cut that we obtain in Lemma 2.1, which is decreasing exponentially with the dimension. Improving this bound, namely obtaining a bound without exponential dependency in the dimension, would yield a “more reasonable” number of levels required to divide the diameter by two in Proposition 1.1.

1.2 $N(X) \longrightarrow +\infty$ in probability

According to Lemma 20.1 in Devroye et al. (1996) and the remark that follows, it is sufficient to show that the number of regions is $o(n)$. For each n , by construction, $T^0(\alpha \log n)$ has $2^{\alpha \log n} = n^{\alpha \log 2}$ leafs. Since $\alpha \log 2 < 1$, $2^{\alpha \log n} = o(n)$ as $n \rightarrow +\infty$. \square

1.3 $\text{diam}(A(X)) \longrightarrow 0$ in probability

Let $0 < \varepsilon < 1$. In this section, we show that

$$\mathbb{P}(\text{diam}(A(X)) > \varepsilon) \longrightarrow 0 \quad \text{when } n \rightarrow +\infty.$$

We first notice that

$$\mathbb{P}(\text{diam}(A(X)) > \varepsilon) \leq \max_i \mathbb{P}(\text{diam}(A_{i,n}) > \varepsilon).$$

Let A be the leaf of $T^0(p_n)$ with maximal diameter and define $\pi := \left\lceil \frac{\log(\sqrt{d}) - \log \varepsilon}{\log 2} \right\rceil$, so that $\varepsilon > \sqrt{d}/2^\pi$. We write

$$\mathbb{P}(\text{diam}(A) > \varepsilon) \leq \mathbb{P}\left(\text{diam}(A) > \frac{\sqrt{d}}{2^\pi}\right).$$

Define C_1, \dots, C_{p_n} the path from $C_0 = [0, 1]^d$ to $C_{p_n} = A$ in the tree $T^0(p_n)$. Set $k = \lfloor \frac{p_n}{\pi} \rfloor$. Set $A^{(0)} = C_0$, $A^{(1)} = C_k$, $A^{(2)} = C_{2k}$, \dots , $A^{(\pi-1)} = C_{(\pi-1)k}$ and $A^{(\pi)} = A$. We define the event $E_j := \{\text{diam}(A^{(j)}) > \sqrt{d}/2^j\}$. Then

$$\begin{aligned} \mathbb{P}\left(\text{diam}(A) > \frac{\sqrt{d}}{2^\pi}\right) &= \mathbb{P}(E_\pi | E_{\pi-1}) \cdot \mathbb{P}(E_{\pi-1}) + \mathbb{P}(E_\pi | E_{\pi-1}^c) \cdot \mathbb{P}(E_{\pi-1}^c) \\ &\hspace{15em} \text{(law of total probability)} \\ &\leq \mathbb{P}(E_\pi | E_{\pi-1}^c) + \mathbb{P}(E_{\pi-1}). \end{aligned}$$

Repeating π times this reasoning, and since $\text{diam}(A^{(0)}) \leq \sqrt{d}$ almost surely, we deduce that

$$\mathbb{P}\left(\text{diam}(A^{(\pi)}) > \varepsilon\right) \leq \sum_{t=1}^{\pi} \mathbb{P}\left(\text{diam}(A^{(t)}) > \frac{\sqrt{d}}{2^t} \mid \text{diam}(A^{(t-1)}) \leq \frac{\sqrt{d}}{2^{t-1}}\right).$$

There are always more than k levels between $A^{(tk)}$ and $A^{((t-1)k)}$ by construction. Hence, according to Proposition 1.1,

$$\mathbb{P}(\text{diam}(A) > \varepsilon) \leq \pi \cdot \frac{N_{f,d}(N_{f,d} + 1)}{2} \cdot \gamma_{f,d}^k.$$

Since $k = O(\log n)$ and $\gamma_{f,d} \in (0, 1)$, we can conclude the proof. \square

2 Auxiliary results

The key in proving Proposition 1.1 is to show that, for a given cell, the probability of a “good cut” is greater than the probability of a “bad cut.” We thus proceed to prove a lower bound for the probability of a good cut (Section 2.1) and an upper bound for the probability of a bad cut (Section 2.2). Since the first cell is the unit cube and all subsequent cells are obtained by intersection with half-spaces, note that any cell of the comparison tree is a full-dimensional convex polytope almost surely. Thus we state and prove our results for such objects.

2.1 Good cuts

The following Lemma is an analogous of Lemma 10 in Dasgupta and Freund (2008). It provides a lower bound on the probability of cleanly separating faraway balls.

Lemma 2.1 (Probability of good cut is lower bounded). *Suppose that Assumption 1.1 holds. Let $A \subset \mathbb{R}^d$ be a full-dimensional convex polytope such that $\text{diam}(A) \leq D < +\infty$. Let $c_r > 10$ be a constant. Pick any two balls $B := \mathcal{B}(z, r)$ and $B' := \mathcal{B}(z', r)$ such that*

- (i) both B and B' intersect A ;
- (ii) their radius is at most D/c_r ;
- (iii) the distance between their centers satisfies $\|z - z'\| \geq D/2 - r$.

Then, if X_1 and X_2 are chosen independently from A according to the distribution of X ,

$$\mathbb{P}(A \cap B \subset A \cap H_{X_1} \text{ and } A \cap B' \subset A \cap H_{X_2}) \geq 2 \frac{f_{\min}}{f_{\max}} \left(\frac{c_r - 10}{4c_r} \right)^{2d}.$$

As a direct consequence, if $c_r > 50$,

$$\mathbb{P}(A \cap B \subset A \cap H_{X_1} \text{ and } A \cap B' \subset A \cap H_{X_2}) \geq \frac{f_{\min}}{f_{\max}} \frac{2}{25^d}.$$

While the statement of Lemma 2.1 is close to that of Lemma 10 in Dasgupta and Freund (2008), a major difference lies in the quality of the bound we obtain. Indeed, our bound depends exponentially in the dimension, therefore becoming arbitrarily loose for large values of d .

Proof. The proof follows the following scheme. First, we conveniently restrict ourselves to the case where the centers of B and B' both belong to A by geometric arguments. We then use Lemma 2.5 to lower bound the probability of a good split by the probability that x and y belong to certain balls γ and γ' . We conclude the proof by finding an upper bound for the volume of A and a lower bound for the volume of $\gamma \cap A$. We refer to Figure 2 throughout this proof.

Preliminary computations. Set $a := \pi_A(z)$, $a' := \pi_A(z')$, $\beta := \mathcal{B}(a, r)$, and $\beta' := \mathcal{B}(a', r)$. Then, according to Lemma 2.3, $A \cap B \subset \beta$ and $A \cap B' \subset \beta'$. For any $x, y \in A$ such that $\beta \subset H_x$ and $\beta' \subset H_y$. Since $A \cap B \subset \beta$, we have $A \cap B \subset H_x$. Furthermore, $A \cap B \subset A$, thus $A \cap B \subset A \cap H_x$. A similar reasoning shows that $A \cap B' \subset A \cap H_y$. Hence

$$\mathbb{P}(A \cap B \subset A \cap H_{X_1} \text{ and } A \cap B' \subset A \cap H_{X_2}) \geq \mathbb{P}(\beta \subset H_{X_1} \text{ and } \beta' \subset H_{X_2}).$$

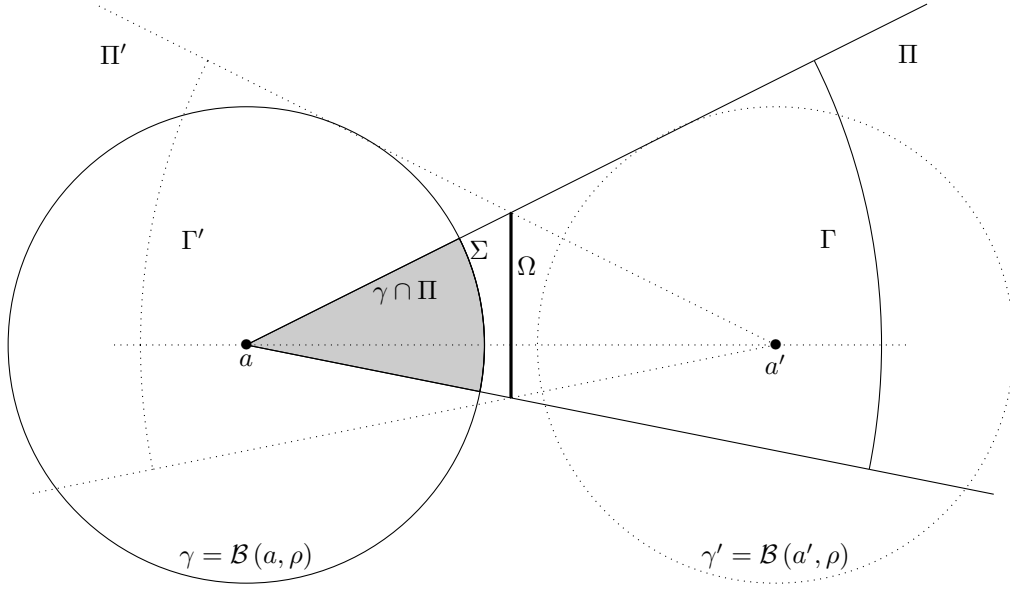


Figure 2: Construction of Ω , Π and Σ . The central thick line represents Ω , the intersection between A and the hyperplane $\Delta(a, a')$. The half-cone Π is the union for all $\omega \in \Omega$ of the half-lines $[a, \omega]$. Finally, the spherical cap Σ is defined as the intersection between $\mathcal{S}(a, \rho)$ and Π . In dotted lines we draw the counter-parts of these objects for a' . The gray area represents $\gamma \cap \Pi$.

Set $\delta := \|a - a'\|$. Since $a \in B$ and $a' \in B'$, by the triangle inequality, $\|a - a'\| \geq \|z - z'\| - 2r$. By hypothesis, $\|z - z'\| \geq D/2 - r$ and $r \leq D/c_r$, thus

$$\|a - a'\| \geq \frac{D}{2} - 3r \geq \frac{c_r - 6}{2c_r} \cdot D.$$

Define $\rho := \|a - a'\|/2 - r$. We have $\rho \geq \frac{c_r - 10}{4c_r} \cdot D$. In particular, as $c_r > 10$, $\rho > 0$. Then, according to Lemma 2.5,

$$\mathbb{P}(\beta \subset H_{X_1} \text{ and } \beta' \subset H_{X_2}) \geq \mathbb{P}(X_1 \in \gamma \text{ and } X_2 \in \gamma' \text{ or } X_2 \in \gamma \text{ and } X_1 \in \gamma'),$$

where $\gamma := \mathcal{B}(a, \rho)$ and $\gamma' := \mathcal{B}(a', \rho)$. Since X_1 and X_2 are independent and identically distributed and $\gamma \cap \gamma' = \emptyset$,

$$\mathbb{P}(X_1 \in \gamma \text{ and } X_2 \in \gamma' \text{ or } X_2 \in \gamma \text{ and } X_1 \in \gamma') \geq 2\mathbb{P}(X_1 \in \gamma)\mathbb{P}(X_2 \in \gamma').$$

Since we sample X_1 and X_2 according to the law of X restricted to A and since Assumption 1.1 holds,

$$\mathbb{P}(X_1 \in \gamma) \geq \frac{f_{\min}}{f_{\max}} \frac{\text{Vol}_d(\gamma \cap A)}{\text{Vol}_d(A)}.$$

In the next paragraphs, we find an upper bound for $\text{Vol}_d(A)$ and a lower bound for $\text{Vol}_d(\gamma \cap A)$. We will see that the latter also holds for γ' .

Upper bound for $\text{Vol}_d(A)$. We refer to Figure 2 for the geometric constructions that follow. Let us first define $\Omega := \Delta(a, a') \cap A$ the intersection between the convex polytope A and the hyperplane $\Delta(a, a')$. We also need to define Π the set of all half-lines going from a through Ω , namely,

$$\Pi := \{a + t(w - a) \mid \omega \in \Omega \text{ and } t > 0\},$$

and the conic section $\Gamma := \mathcal{B}(a, \text{diam}(A)) \cap \Pi$. We claim that $A \cap H_{a'} \subset \Gamma$. Indeed, let $\xi \in A \cap H_{a'}$. Since $\xi \in H_{a'}$, $[a, \xi]$ intersects $\Delta(a, a')$ in a unique point, say ζ . By convexity, the segment $[a, \xi]$ is contained into A . In particular, $\zeta \in A$. Thus $\zeta \in \Delta(a, a') \cap A = \Omega$, and

$$\xi = a + \frac{\|\xi - a\|}{\|\zeta - a\|}(\zeta - a) \in A.$$

Moreover, since $\xi \in A$,

$$\|a - \xi\| \leq \sup_{s \in A} \|a - s\| = \text{diam}(A),$$

and $\xi \in \mathcal{B}(a, \text{diam}(A))$. A similar reasoning shows that $A \cap H_a \subset \Gamma'$, where Γ' is the symmetric of Γ with respect to $\Delta(a, a')$. Therefore,

$$\text{Vol}_d(A) \leq 2 \text{Vol}_d(\Gamma).$$

Define the hyperspherical cap $\Sigma := \mathcal{S}(a, \rho) \cap \Pi$. Then we can express the volume of the conic section Γ as

$$\text{Vol}_d(\Gamma) = \frac{\text{Vol}_{d-1}(\Sigma)}{\text{Vol}_{d-1}(\mathcal{S}(a, \rho))} \text{Vol}_d(\mathcal{B}(a, \text{diam}(A))),$$

which leads to

$$\text{Vol}_d(A) \leq \frac{2 \text{Vol}_{d-1}(\Sigma)}{\text{Vol}_{d-1}(\mathcal{S}(a, \rho))} \text{Vol}_d(\mathcal{B}(a, \text{diam}(A))). \quad (2.2)$$

Lower bound for $\text{Vol}_d(\gamma \cap A)$. By convexity, $\gamma \cap \Pi \subset \gamma \cap A$. Moreover,

$$\text{Vol}_d(\gamma \cap \Pi) = \frac{\text{Vol}_{d-1}(\Sigma)}{\text{Vol}_{d-1}(\mathcal{S}(a, \rho))} \text{Vol}_d(\mathcal{B}(a, \rho)).$$

Hence the following lower bound holds:

$$\text{Vol}_d(\gamma \cap A) \geq \frac{\text{Vol}_{d-1}(\Sigma)}{\text{Vol}_{d-1}(\mathcal{S}(a, \rho))} \text{Vol}_d(\mathcal{B}(a, \rho)). \quad (2.3)$$

Conclusion. Putting together Eq. (2.2) and (2.3), we obtain

$$\mathbb{P}(X_1 \in \gamma) \geq \frac{f_{\min}}{f_{\max}} \frac{\text{Vol}_d(\mathcal{B}(a, \rho))}{\text{Vol}_d(\mathcal{B}(a, \text{diam}(A)))} = \frac{f_{\min}}{f_{\max}} \left(\frac{\rho}{\text{diam}(A)} \right)^d.$$

Since $\rho \geq (c_r - 10)/(4Dc_r)$ and $\text{diam}(A) \leq D$, we deduce that

$$\mathbb{P}(X_1 \in \gamma) \geq \frac{f_{\min}}{f_{\max}} \left(\frac{c_r - 10}{4c_r} \right)^d.$$

We conclude the proof by using the preliminary computations. \square

2.2 Bad cuts

We now focus on the probability of a “bad split,” that is, $\Delta(x, y)$ intersects both $\mathcal{B}(z, r)$ and $\mathcal{B}(z', r)$. The following result is an analogous of Lemma 11 in Dasgupta and Freund (2008).

Lemma 2.2 (Probability of bad cut is upper bounded). *Suppose that assumption 1.1 holds. Let $A \subset \mathbb{R}^d$ be a full-dimensional convex polytope such that $\text{diam}(A) \leq D < +\infty$. Let $c_r > 10$ be a constant. Pick any two balls $B := \mathcal{B}(z, r)$ and $B' := \mathcal{B}(z', r)$ such that*

- (i) both B and B' intersect A ;
- (ii) their radius is at most D/c_r ;
- (iii) the distance between their centers satisfies $\|z - z'\| \geq D/2 - r$.

Then, if X_1 and X_2 are chosen independently from A according to the distribution of X ,

$$\mathbb{P}(A \cap B \cap \Delta(X_1, X_2) \neq \emptyset \text{ and } A \cap B' \cap \Delta(X_1, X_2) \neq \emptyset) \leq \frac{f_{\max}}{f_{\min}} \frac{32dc_r}{(c_r - 2)(c_r - 6)}.$$

As a direct consequence, if $c_r > 15$,

$$\mathbb{P}(A \cap B \cap \Delta(X_1, X_2) \neq \emptyset \text{ and } A \cap B' \cap \Delta(X_1, X_2) \neq \emptyset) \leq \frac{f_{\max}}{f_{\min}} \frac{64d}{c_r}.$$

Note that, as in Lemma 2.1, the bound we obtain worsens as the dimension increases.

Proof. We first restrict ourselves to the case where the centers of B and B' both belong to A with the same argument than in the proof of Lemma 2.1. Namely, define $a := \pi_A(z)$, $a' := \pi_A(z')$, $\beta := \mathcal{B}(a, r)$, $\beta' := \mathcal{B}(a', r)$. According to Lemma 2.3, $A \cap B \subset \beta$ and $A \cap B' \subset \beta'$. Thus

$$\begin{aligned} & \mathbb{P}(A \cap B \cap \Delta(X_1, X_2) \neq \emptyset \text{ and } A \cap B' \cap \Delta(X_1, X_2) \neq \emptyset) \\ & \leq \mathbb{P}(\beta \cap \Delta(X_1, X_2) \neq \emptyset \text{ and } \beta' \cap \Delta(X_1, X_2) \neq \emptyset). \end{aligned}$$

For any $x \in \mathbb{R}^d$, define B_x the set of points y such that $\Delta(x, y)$ is a bad cut, that is,

$$B_x := \{y \in \mathbb{R}^d \mid \beta \cap \Delta(x, y) \neq \emptyset \text{ and } \beta' \cap \Delta(x, y) \neq \emptyset\}.$$

Then, since we sample X_1 according to the law of X restricted to A and since we assume Assumption 1.1 to be true,

$$\mathbb{P}(\beta \cap \Delta(X_1, X_2) \neq \emptyset \text{ and } \beta' \cap \Delta(X_1, X_2) \neq \emptyset) \leq \frac{f_{\max}}{f_{\min}} \frac{\mathbb{E}[\text{Vol}_d(B_{X_1} \cap A)]}{\text{Vol}_d(A)},$$

where the expectation is relative to the random variable X_1 .

Upper bound for $\text{Vol}_d(B_x \cap A)$. Let $x \in A$ and $y \in B_x$. By Lemma 2.4,

$$\begin{cases} (\|x - a\| - 2r)^+ & \leq \|y - a\| \leq \|x - a\| + 2r \\ (\|x - a'\| - 2r)^+ & \leq \|y - a'\| \leq \|x - a'\| + 2r. \end{cases}$$

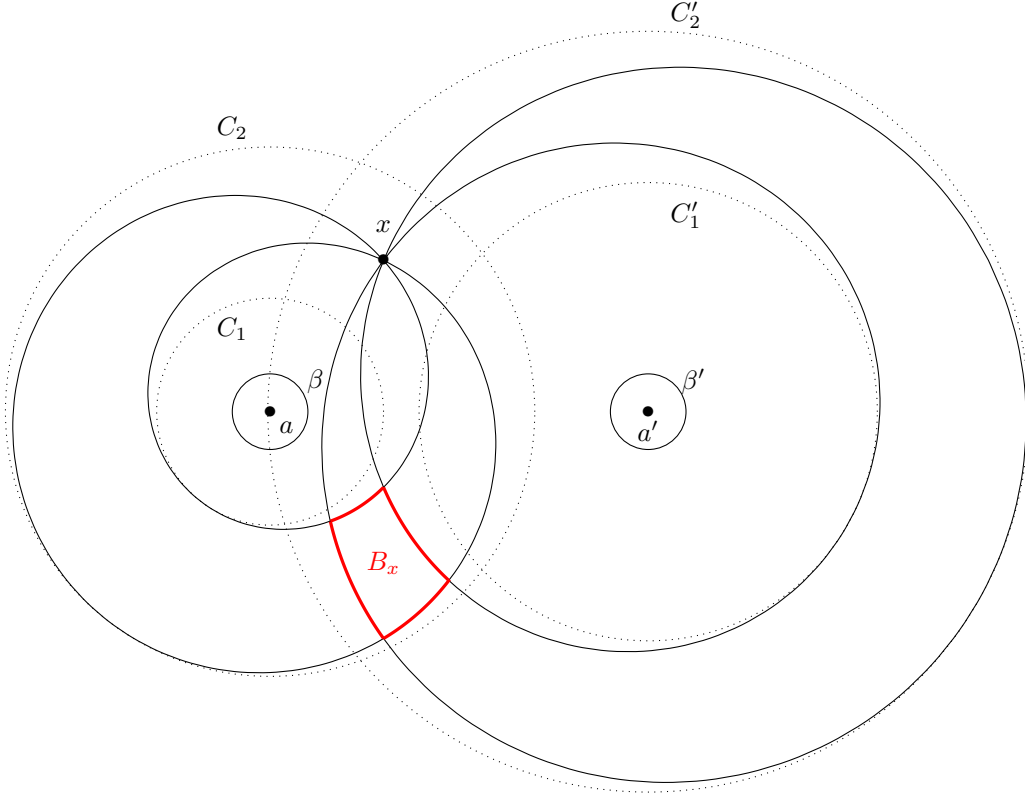


Figure 3: Sketch of B_x in \mathbb{R}^2 . For a fixed x , B_x is the set of all y such that $\Delta(x, y)$ cuts both β and β' (border marked in red). We show that B_x is the intersection of two geometric loci (solid lines border). In particular, B_x is included in the intersection of two annuli (border in dotted lines).

Equivalently, $B_x \subset \mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$, where we defined $r_1 := \|x - a\| - 2r$, $r_2 := \|x - a\| + 2r$, $r'_1 := \|x - a'\| - 2r$ and $r'_2 := \|x - a'\| + 2r$. Recall that $\mathcal{A}(a, r_1, r_2) = \mathcal{B}(a, r_2)$ whenever $r_1 \leq 0$. See Figure 3 for an illustration.

For any $\xi \in (a, a')$, denote by D_ξ the hyperplane orthogonal to (a, a') and passing through ξ . According to Lemma 2.6, the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ along the (a, a') axis is upper bounded by $16D/(c_r - 2)$, hence there exists ξ^- and $\xi^+ \in (a, a')$ such that $\|\xi^+ - \xi^-\| \leq 16D/(c_r - 2)$ and $B_x \cap A$ is contained between D_{ξ^-} and D_{ξ^+} . For each $\xi \in (a, a')$, set $\Omega_\xi := D_\xi \cap A$. There exists $\xi^* \in [\xi^-, \xi^+]$ such that $\text{Vol}_{d-1}(\Omega_{\xi^*})$ is maximal, and

$$\begin{aligned}
 \text{Vol}_d(B_x \cap A) &= \int_{\xi \in [\xi^-, \xi^+]} \text{Vol}_{d-1}(\Omega_\xi) \, d\xi \\
 &\leq \|\xi^+ - \xi^-\| \cdot \text{Vol}_{d-1}(\Omega_{\xi^*}) \\
 \text{Vol}_d(B_x \cap A) &\leq \frac{16}{c_r - 2} \cdot D \text{Vol}_{d-1}(\Omega_{\xi^*}). \tag{2.4}
 \end{aligned}$$

Lower bound for $\text{Vol}_d(A)$. Suppose that ξ^* belongs to the segment $[a, a']$. By convexity, A contains the (disjoint) union of the two hyperpyramids of apexes a and a' with $(d-1)$ -dimensional basis Ω_{ξ^*} , which we denote by Q and Q' . Therefore,

$$\begin{aligned} \text{Vol}_d(A) &\geq \text{Vol}_d(Q) + \text{Vol}_d(Q') \\ &= \frac{\|a - \xi^*\| \text{Vol}_{d-1}(\Omega_{\xi^*})}{d} + \frac{\|a' - \xi^*\| \text{Vol}_{d-1}(\Omega_{\xi^*})}{d} \\ &= \frac{\delta \text{Vol}_{d-1}(\Omega_{\xi^*})}{d}. \end{aligned}$$

Since $\delta \geq (c_r - 6)D/(2c_r)$,

$$\text{Vol}_d(A) \geq \frac{c_r - 6}{2dc_r} \cdot D \text{Vol}_{d-1}(\Omega_{\xi^*}). \quad (2.5)$$

A similar reasoning holds whenever ξ^* does not belong to $[a, a']$.

Conclusion. Putting together Eq. (2.4) and (2.5), we obtain

$$\mathbb{P}(\beta \cap \Delta(X_1, X_2) \neq \emptyset \text{ and } \beta' \cap \Delta(X_1, X_2) \neq \emptyset) \leq \frac{f_{\max}}{f_{\min}} \frac{32dc_r}{(c_r - 2)(c_r - 6)},$$

which concludes the proof. \square

Note that in the plane defined by a , a' and x , we can actually describe precisely the shape of the curves defining the border of B_x —see Figure 3. These curves correspond to the images of x by all the symmetries with respect to a line tangent to β or β' . Individually, they are called the *orthotomics of a circle*, or second caustic (Lawrence, 2013, p. 60).

2.3 Technical results

This first lemma is used in the proofs of Lemma 2.1 and 2.2 to deal with cases where the center of B or B' does not belong to A . See Figure 4 for an illustration of such a situation.

Lemma 2.3 (Construction of β). *Let $A \subset \mathbb{R}^d$ be a convex compact set and $\mathcal{B}(z, r)$ be a ball that intersects A . Define $\beta := \mathcal{B}(\pi_A(z), r)$. Then $A \cap B \subset \beta$.*

Proof. Set $a := \pi_A(z)$. Let x be an element of $A \cap B$. Then,

$$\begin{aligned} \|x - a\|^2 &= \langle x - a, x - a \rangle \\ &= \langle x - z + z - a, x - z + z - a \rangle \\ &= \|x - z\|^2 + 2\langle x - z, z - a \rangle + \|z - a\|^2 \\ \|x - a\|^2 &= \|x - z\|^2 + 2\langle x - a, z - a \rangle - \|z - a\|^2. \end{aligned}$$

Since π_A is the orthogonal projection, given that $x \in A$, we have $\langle x - a, z - a \rangle \leq 0$. Moreover, $\|z - a\| \geq 0$, thus $\|x - a\|^2 \leq \|x - z\|^2$. But x also belongs to B , hence $\|x - z\| \leq r$. As a consequence, $\|x - a\| \leq r$, that is, $x \in \beta$. \square

The next lemma shows that, for a given x , the set of every possible y such that $\Delta(x, y)$ intersects $\mathcal{B}(a, r)$ is contained into an annulus centered in a . We refer to Figure 5 for an illustration.

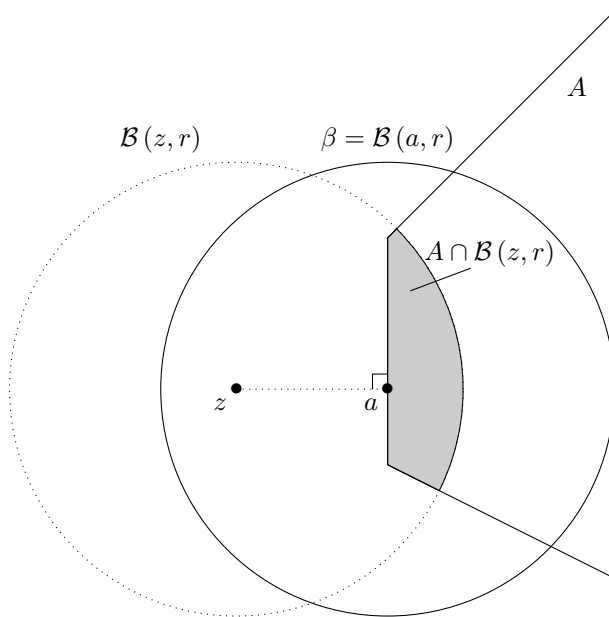


Figure 4: Construction of β . The point a is the image of z by the orthogonal projection on A . The ball β has the same radius than $\mathcal{B}(z, r)$ and contains $A \cap \mathcal{B}(z, r)$, which is marked in gray.

Lemma 2.4 (Localization of B_x). *Let $a, x \in \mathbb{R}^d$ and $r > 0$. Then, for any $y \in \mathbb{R}^d$ such that $\Delta(x, y) \cap \mathcal{B}(a, r)$ is non-empty,*

$$(\|x - a\| - 2r)^+ \leq \|y - a\| \leq \|x - a\| + 2r.$$

Proof. Let $y \in \mathbb{R}^d$ such that $\Delta(x, y) \cap \mathcal{B}(a, r)$ is non-empty. In particular, there exists $b \in \mathbb{R}^d$ such that $\|y - b\| = \|x - b\|$ and $\|a - b\| \leq r$. By the triangle inequality,

$$\left| \|y - a\| - \|y - b\| \right| \leq \|a - b\| \leq r.$$

Hence

$$\begin{cases} \|y - a\| \leq r + \|y - b\| = r + \|x - b\| \\ \|y - a\| \geq -r + \|y - b\| = -r + \|x - b\|. \end{cases}$$

Since $\left| \|x - b\| - \|a - b\| \right| \leq \|x - a\|$ (again by the triangle inequality), we have

$$\begin{cases} \|y - a\| \leq \|x - a\| + 2r \\ \|y - a\| \geq \|x - a\| - 2r. \end{cases}$$

□

We now present a result stating that, for any two points $a, a' \in \mathbb{R}^d$, there exists a simple set of possible x and y such that balls with center a and a' are well-separated by $\Delta(x, y)$. It is the key element in the proof of Lemma 2.1.

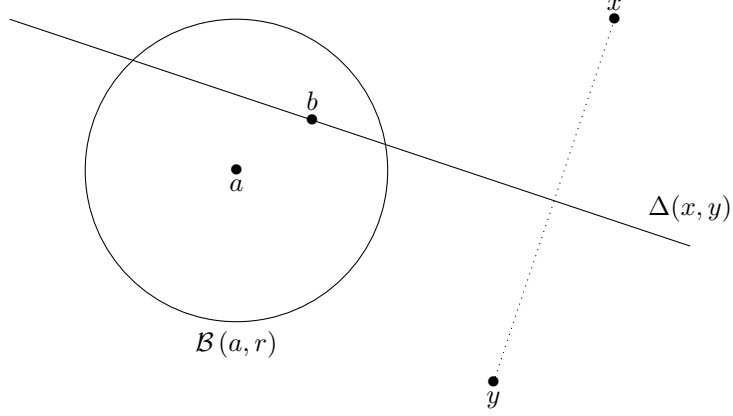


Figure 5: B_x is included in the intersection of two annuli. As in the proof of Lemma 2.4, a and x are fixed, and y is such that $\Delta(x, y)$ intersects $\mathcal{B}(a, r)$. Then y belongs to an annulus of radii $(\|x - a\| - 2r)^+$ and $\|x - a\| + 2r$.

Lemma 2.5 (Sufficient condition for a good cut). *Let $a, a' \in \mathbb{R}^d$. Let $0 < r < \|a - a'\|/2$ and set $\rho := \|a - a'\|/2 - r$. Then, for any $x \in \mathcal{B}(a, \rho)$ and $y \in \mathcal{B}(a', \rho)$, we have $\mathcal{B}(a, r) \subset H_x$ and $\mathcal{B}(a', r) \subset H_y$.*

Remark 2.1. Note that Lemma 2.5 holds in any metric space (\mathcal{X}, δ) since the proof only uses the triangle inequality.

Proof. We refer to Figure 6 for this proof. We have to prove that for any $s \in \mathcal{B}(a, r)$, $\delta(s, x) \leq \delta(s, y)$ (the case $t \in \mathcal{B}(a', r)$ is identical up to notations). We first write

$$\delta(s, x) \leq \delta(s, a) + \delta(a, x) \leq r + \rho = \delta(a, a')/2,$$

where we used (i) the triangle inequality, (ii) $s \in \mathcal{B}(a, r)$ and $x \in \mathcal{B}(a, \rho)$, (iii) the definition of ρ . Then,

$$\delta(a, a') \leq \delta(a, y) + \delta(a', y) \leq \delta(a, y) + \rho,$$

where we used (i) triangle inequality, (ii) $y \in \mathcal{B}(a', \rho)$. Thus $\delta(a, y) \geq \delta(a, a') - \rho$. Moreover,

$$\delta(a, y) \leq \delta(a, s) + \delta(s, y) \leq r + \delta(s, y),$$

where we used (i) triangle inequality, (ii) $s \in \mathcal{B}(a, r)$. Combining the two, we get

$$\delta(s, y) \geq \delta(a, a') - (r + \rho) = \delta(a, a')/2.$$

Therefore $\delta(s, y) \geq \delta(s, x)$ and we can conclude. \square

Finally, we state and prove a technical lemma used in the proof of Lemma 2.2 to control the size of the intersection of two annuli.

Lemma 2.6 (B_x has small width). *Assume the set of hypotheses of Lemma 2.2 and define r_1, r_2, r'_1 and r'_2 as in the proof of Lemma 2.2. Then there exist two hyperplanes L_x*

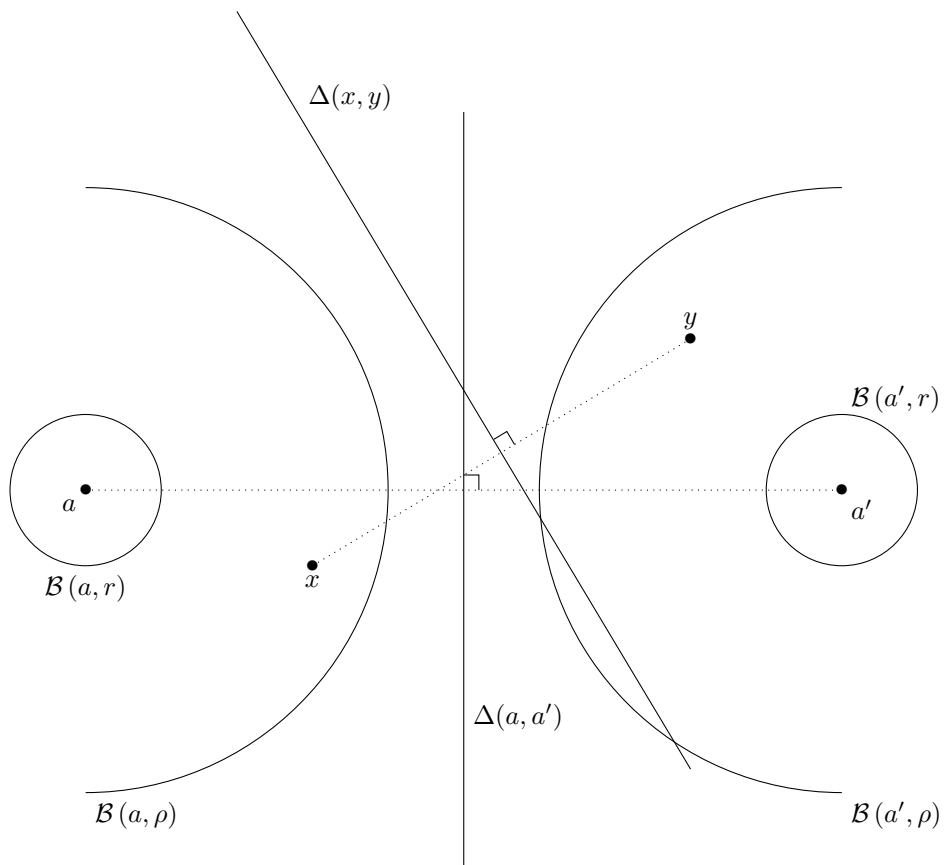


Figure 6: Guaranteed good cut. Set $a, a' \in \mathbb{R}^d$ and $\rho = \|a - a'\|/2 - r$. Then, for any $x \in \mathcal{B}(a, \rho)$ and $y \in \mathcal{B}(a', \rho)$, the hyperplane $\Delta(x, y)$ separates cleanly $\mathcal{B}(a, r)$ from $\mathcal{B}(a', r)$.

and L'_x , orthogonal to (a, a') , such that the intersection of $\mathcal{A}(a, r_1, r_2)$ and $\mathcal{A}(a', r'_1, r'_2)$ is included between L_x and L'_x . Additionally,

$$\delta(L_x, L'_x) \leq \frac{16D}{c_r - 2}. \quad (2.6)$$

Even though the statement of Lemma 2.6 may seem intuitive at first sight (since the intersection is contained in two annuli of width $O(D/c_r)$, one would expect its width to be of the same order), we do not know of a simpler proof. We believe that it is necessary to describe precisely the intersection of the two annuli depending on the radii in order to make sure that the situation where the two annuli are overlapping is excluded. Indeed, in this case the width of the intersection is *not* bounded by a quantity depending on D/c_r but rather on D , since it has the same order than the diameter of the annuli.

Proof. By rotational symmetry around (a, a') , it suffices to prove the result in a 2-plane containing a and a' . Hence from now on we work in the plane P defined by the triple (x, a, a') . We first describe the shape of the intersection between the two annuli depending on the position of x relatively to a and a' . Then, in each case, we bound the width of the intersection in the direction of the (a, a') axis.

Shape of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$. Let us equip P with an orthogonal frame such that $a = (0, 0)$, $a' = (\delta, 0)$ and $x = (x_1, x_2)$. The width of the intersection is invariant by line symmetry with respect to $\Delta(a, a')$ and (a, a') , thus we can restrict our analysis to the quadrant defined by $x_1 \leq \delta/2$ and $x_2 > 0$. In particular, $\|x - a\| \leq \|x - a'\|$. Define $C_i := \mathcal{C}(a, r_i)$ and $C'_i := \mathcal{C}(a', r'_i)$ for $i \in \{1, 2\}$. The shape of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ depends on the mutual intersections between C_1, C_2, C'_1 and C'_2 . Recall that $\mathcal{C}(a, \rho) \cap \mathcal{C}(a', \rho) \neq \emptyset$ if, and only if,

$$|\rho - \rho'| \leq \delta \leq \rho + \rho'.$$

We now proceed to describe these intersection depending on the position of x relatively to a and a' .

- Since $r > 0$, $r_1 < r_2$ and $r'_1 < r'_2$ and thus $C_1 \cap C_2 = C'_1 \cap C'_2 = \emptyset$;
- By the triangle inequality, $|r_2 - r'_2| = ||x - a| - |x - a'|| \leq \delta$ and $r_2 + r'_2 = \|x - a\| + \|x - a'\| + 4r \geq \delta$, hence $C_2 \cap C'_2$ is always non-empty;
- By the triangle inequality, $|r_1 - r'_1| = ||x - a| - |x - a'|| \leq \delta$. Hence $C_1 \cap C'_1$ is non-empty if, and only if, $r_1 + r'_1 \geq \delta$, that is, $\|x - a\| + \|x - a'\| \geq \delta - 4r$. The border is an ellipse with focal points a, a' and semi-major axis $(\delta + 4r)/2$.
- By the triangle inequality, $r_1 + r'_2 = \|x - a\| + \|x - a'\| \leq \delta$. Since $\|x - a\| \leq \|x - a'\|$, $|r_1 - r'_2| = 4r - \|x - a\| + \|x - a'\|$. Thus $C_1 \cap C'_2$ is non-empty if, and only if, $\|x - a'\| - \|x - a\| \leq \delta - 4r$. The border is a hyperbola with focal points a, a' and semi-major axis $(\delta - 4r)/2$.
- By the triangle inequality, $r'_1 + r_2 = \|x - a'\| + \|x - a\| \geq \delta$. Moreover, $|r'_1 - r_2| = ||x - a'\| - \|x - a\| - 4r|$. Again, the triangle inequality yields $\|x - a'\| - \|x - a\| \leq \delta \leq \delta + 4r$. On the other side, $\|x - a\| - \|x - a'\| \leq 0 \leq \delta - 4r$ because $r < \delta/4$. Hence $C_2 \cap C'_1$ is always non-empty.

The different cases are summarized in Figure 7, and we provide a visual depiction of the intersection for each case in Figure 8. Note that in Case III, $\|x - a\| \leq 2r$ is a possibility, implying $r_1 < 0$. In this event, we see in Figure 8 that the extremal points are the same.

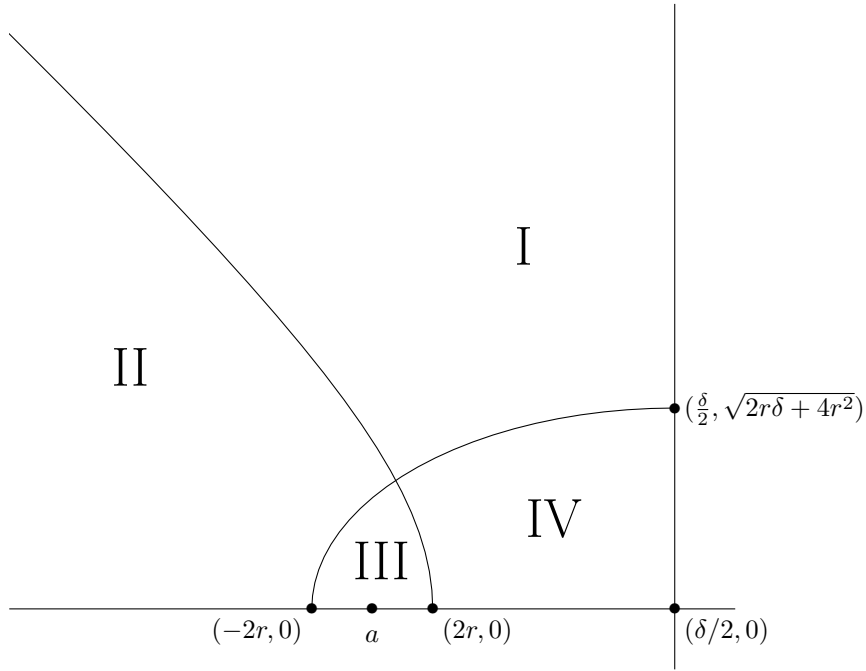


Figure 7: Shape of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ (I). Depending on the relative position of x with respect to a and a' , the shape of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ changes. *Case I*: $C_1 \cap C'_1$ and $C_1 \cap C'_2$ are both non-empty. *Case II*: $C_1 \cap C'_1$ is non-empty, whereas $C_1 \cap C'_2$ is. *Case III*: $C_1 \cap C'_1$ and $C_1 \cap C'_2$ are both empty. *Case IV*: $C_1 \cap C'_2$ is non-empty whereas $C_1 \cap C'_1$ is. The shape of B_x as well as $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ in this last case is depicted in Figure 3.

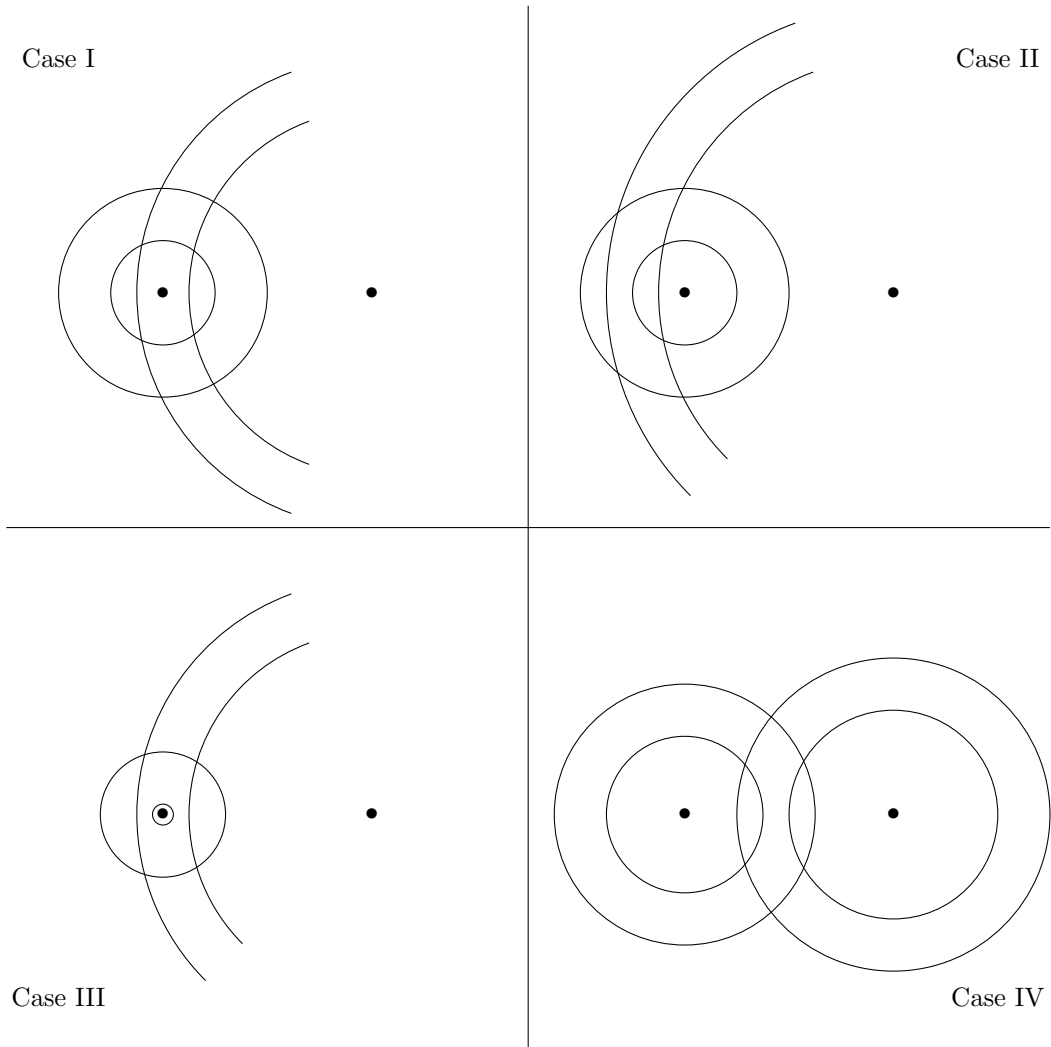


Figure 8: Shape of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ (II). For each case described in Figure 7, we sketch $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$. Note that the points realizing the minimum and maximum abscissa in each case are different, leading to different bounds on the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$.

Width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$. For each case, we show that Eq. (2.6) holds. Recall that we assumed $r/\delta \leq 2/(c_r - 2)$ and $\text{diam}(A) \leq D$. We will use the fact that

$$\frac{\|x - a\|^2 - \|x - a'\|^2}{2\delta} = \frac{x_1^2 + x_2^2 - x_1'^2 + 2\delta x_1 + \delta^2 + x_2^2}{2\delta} = x_1 - \frac{\delta}{2}.$$

- Case I: The left-most points of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ belong to $C_1 \cap C'_2$. We solve

$$\begin{cases} \xi_1^2 + \xi_2^2 & = r_1^2 = (\|x - a\| - 2r)^2 \\ (\xi_1 - \delta)^2 + \xi_2^2 & = r_2'^2 = (\|x - a'\| + 2r)^2. \end{cases}$$

and find

$$\xi_1 = x_1 - \frac{2r}{\delta} (\|x - a\| + \|x - a'\|).$$

The right-most points of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ belong to $C_2 \cap C'_1$. We solve

$$\begin{cases} \zeta_1^2 + \zeta_2^2 & = r_2^2 = (\|x - a\| + 2r)^2 \\ (\zeta_1 - \delta)^2 + \zeta_2^2 & = r_1'^2 = (\|x - a'\| - 2r)^2. \end{cases}$$

and find

$$\zeta_1 = x_1 + \frac{2r}{\delta} (\|x - a\| + \|x - a'\|).$$

Thus the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ along (a, a') is given by

$$|\zeta_1 - \xi_1| = \frac{4r}{\delta} (\|x - a\| + \|x - a'\|) \leq \frac{16D}{c_r - 2}.$$

- Case II: The left-most point of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ belongs to $C'_2 \cap (a, a')$, and we have

$$\xi_1 = \delta - r'_2 = \delta - \|x - a'\| - 2r.$$

The right-most points belongs to $C_2 \cap C'_1$, and we have, as in Case I,

$$\zeta_1 = x_1 + \frac{2r (\|x - a\| + \|x - a'\|)}{\delta}.$$

Thus the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ along (a, a') is given by

$$|\zeta_1 - \xi_1| = \frac{2r}{\delta} (\|x - a\| + \|x - a'\|) + \|x - a'\| + x_1 - \delta - 2r.$$

The equation of the asymptotes of the hyperbola $\|x - a'\| - \|x - a\| = \delta - 4r$ are given by

$$x_2 = \pm \frac{2\sqrt{2r\delta - 4r^2}}{\delta - 4r} (x_1 - \delta/2),$$

and considering the lines parallel to these asymptotes passing through $(\delta, 0)$ we deduce that, in case II,

$$\frac{x_2^2}{(x_1 - \delta)^2} \leq \frac{4(2r\delta - 4r^2)}{(\delta - 4r)^2} \leq \frac{8r\delta}{(\delta - 4r)^2} \leq 8\frac{r}{\delta} \frac{1}{(1 - 4\frac{r}{\delta})^2} \leq \frac{16(c_r - 10)^2}{(c_r - 2)^3}.$$

Thus

$$\|x - a'\| = \sqrt{(x_1 - \delta)^2 + x_2^2} \leq |x_1 - \delta| \sqrt{1 + \frac{16(c_r - 10)^2}{(c_r - 2)^3}},$$

and we have

$$\|x - a'\| + x_1 - \delta \leq |x_1 - \delta| \left(\sqrt{1 + \frac{16(c_r - 10)^2}{(c_r - 2)^3}} - 1 \right) \leq \frac{8D(c_r - 10)^2}{(c_r - 2)^3},$$

where we used $\sqrt{1 + x^2} - 1 \leq x/2$ in the last inequality. Finally,

$$|\zeta_1 - \xi_1| \leq \frac{8D(c_r - 10)^2}{(c_r - 2)^3} + \frac{8D}{c_r - 2}.$$

- Case III: The left-most point of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ belongs to $C_2 \cap C'_2$. We solve

$$\begin{cases} \xi_1^2 + \xi_2^2 & = r_2^2 = (\|x - a\| + 2r)^2 \\ (\xi_1 - \delta)^2 + \xi_2^2 & = r_2'^2 = (\|x - a'\| + 2r)^2, \end{cases}$$

which yields

$$\xi_1 = x_1 + \frac{2r}{\delta} (\|x - a\| - \|x - a'\|).$$

The right-most points belongs to $C_2 \cap C'_1$, and we have, as in Case I,

$$\zeta_1 = x_1 + \frac{2r(\|x - a\| + \|x - a'\|)}{\delta}.$$

Thus the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ along (a, a') is given by

$$|\zeta_1 - \xi_1| = \frac{4r}{\delta} \|x - a'\| \leq \frac{8D}{c_r - 2}.$$

- Case IV: as in Case I, the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ is given by

$$|\xi_1 - \zeta_1| = \frac{4r(\|x - a\| + \|x - a'\|)}{\delta}.$$

Since in this case $\|x - a\| + \|x - a'\| \leq \delta - 4r$, we have

$$|\xi_1 - \zeta_1| \leq 4r \leq \frac{4D}{c_r}.$$

Overall, since $c_r > 10$, we have shown that the width of $\mathcal{A}(a, r_1, r_2) \cap \mathcal{A}(a', r'_1, r'_2)$ along (a, a') is upper bounded by $16D/(c_r - 2)$. \square

3 Additional experiments

3.1 Comparison of embedding methods

Here we report the results of the comparison between embedding methods. The results are provided as a supplement to Section 4.3. We use a subsample of $n = 500$ points

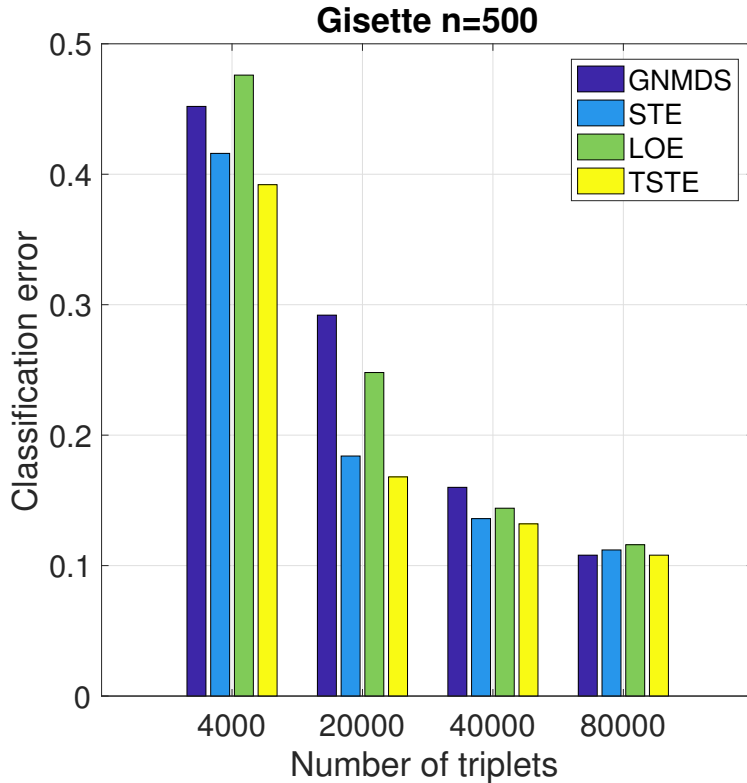


Figure 9: Classification error of the various embedding methods and the KNN algorithm.

from the Gisette dataset (half as training set and half as test set). The dimension $d \in \{10, 20, 30, 40, 50\}$ and $k \in \{2, 6, 10, 14, 18\}$ of the KNN are adjusted with 2-fold cross-validation on the training set. Figure 9 shows the classification error of the four embedding methods: GNMDS (Agarwal et al., 2007), LOE (Terada and von Luxburg, 2014) and STE/TSTE (van der Maaten and Weinberger, 2012).

The TSTE consistently outperforms other methods. Therefore, we use it as the main competitor against our proposed random forest.

3.2 CompRF and subsampling

In this section we investigate the role of subsampling in the performance of the CompRF. To construct each tree of the CompRF, we randomly pick a subsample of $r|S|$ points among the set of training points (S) without replacement and make the tree only based on the subsample. We use the following range for the subsampling ratio: $r \in \{0.1, 0.2, 0.4, 1\}$. The left panel of Figure 10 shows the average classification error of the CompRF for various values of r . The right plot in this figure shows the normalized average MSE of the CompRF for regression datasets. Note that the range of MSE depends on the dataset. To make a unified figure, for each dataset, we divided all average values of the MSE by the maximum value of the MSE on that particular dataset.

Our results hardly show any significant positive effect of subsampling. On the contrary, in

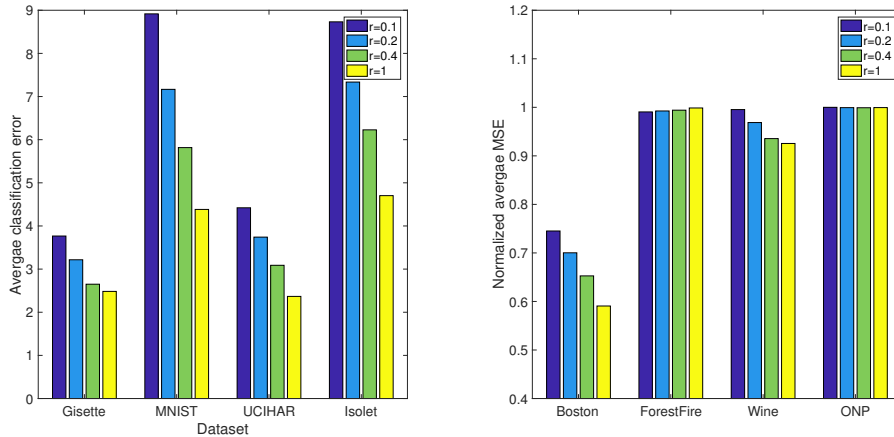


Figure 10: (Left) Average classification error of the ComprRF algorithm on four classification datasets and various subsampling ratios (r). (Right) Normalized average MSE of the ComprRF algorithm on four regression datasets and various subsampling ratios (r). The X-axis denotes the datasets. Note that for each dataset we divided all MSE values by the maximum value of the dataset. In this way bars can be plotted together.

classification tasks we see a significant decrease in error when the whole dataset is used. Only in case of ForestFire dataset do we see some slight improvement.

3.3 Running time of ComprRF vs. Embedding procedures

Here we report the running time of ComprRF in comparison with TSTE embedding combined with KNN. Note that if we apply CART forest after embedding, it can be even more time consuming. In addition, the running time of embedding does not change significantly if we apply the same triplets as the ComprRF or a random subsample of triplets, therefore we report the running time based on the same triplets as the ComprRF.

We use the subsample of Gisette dataset with $n = 1000$ point, similar to the Section 4.3. We perform the embedding with $d = 10$ and $d = 50$ dimensions and fixed $k = 5$. Table 1 shows the running time of the experiments. Since the running time of embedding can change significantly based on the initial conditions, we run embedding algorithms five times and we report the average running time. The algorithms are implemented on a single core CPU and the running times are reported in seconds.

Table 1: Comparison of computation time between ComprRF and TSTE+KNN. The reported values are in seconds.

Number of trees (M)	1	5	10	20
ComprRF	1	4	8	16
TSTE+KNN (d=10)	148	236	350	595
TSTE+KNN (d=50)	185	654	1214	2398

The required running time for the embedding algorithm is orders of magnitude longer than

the CompRF. Moreover, the embedding algorithms need a cross-validation step to adjust the number of dimensions and other parameters of the classifier.

References

- S. Agarwal, J. Wills, L. Cayton, G. Lanckriet, D. Kriegman, and S. Belongie. Generalized non-metric multidimensional scaling. In *AISTATS*, pages 11–18, 2007.
- S. Dasgupta and Y. Freund. Random projection trees and low dimensional manifolds. In *STOC*, pages 537–546, 2008.
- L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*. Springer, 1996.
- J. Lawrence. *A catalog of special plane curves*. Courier Corporation, 2013.
- S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- Y. Terada and U. von Luxburg. Local ordinal embedding. In *ICML*, pages 847–855, 2014.
- L. van der Maaten and K. Weinberger. Stochastic triplet embedding. In *Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2012. Code available at <http://homepage.tudelft.nl/19j49/ste>.