IMPROVED PERFORMANCE GUARANTEES FOR TUKEY'S MEDIAN

STANISLAV MINSKER AND YINAN SHEN

Abstract. Is there a natural way to order data in dimension greater than one? The approach based on the notion of data depth, often associated with the name of John Tukey, is among the most popular. Tukey's depth has found applications in robust statistics, graph theory, and the study of elections and social choice. We present improved performance guarantees for empirical Tukey's median, a deepest point associated with the given sample, when the data-generating distribution is elliptically symmetric and possibly anisotropic. Some of our results remain valid in the wider class of affine equivariant estimators. As a corollary of our bounds, we show that the typical diameter of the set of all empirical Tukey's medians scales like $o(n^{-1/2})$ where n is the sample size. Moreover, when the data are 2-dimensional, we prove that with high probability, the diameter is of order $O(n^{-3/4}\log^{3/2}(n))$.

1. Introduction

The fundamental notions of order statistics and ranks form a basis of many inferential procedures. They are the backbone of methods in robust statistics that studies stability of algorithms under perturbations of data (Huber and Ronchetti, 2009). As the Euclidean space of dimension 2 and higher lacks the canonical order, the "canonical" versions of order statistics and ranks are difficult to construct. Many useful versions have been proposed over the years: examples include the componentwise ranks (Hodges, 1955), spatial ranks and quantiles (Brown, 1983; Koltchinskii, 1997), Mahalanobis ranks (Hallin and Paindaveine, 2002), and recently introduced Monge-Kantorovich ranks and quantiles (Chernozhukov et al., 2017). In this paper, we will focus on the notion of *data depth* that gives rise to the depth-based ranks and quantiles, see (Serfling, 2002) for an overview. The idea of depth and the associated ordering with respect to a probability measure P on $\mathbb{R}$ goes back to Hotelling (1929): given $z \in \mathbb{R}$, its *depth* is defined as the minimum among $P(-\infty, z]$ and $P[z, \infty)$. Hodges (1955) and Tukey (1975) extended this idea to $\mathbb{R}^2$, and finally Donoho (1982); Donoho and Gasko (1992) formalized the general notion of depth of a point $z \in \mathbb{R}^d$ with respect to a probability measure P as the infimum of all univariate depth evaluated over projections of P on the lines passing through z. Equivalently, Tukey's (or half-space) depth of z is

$$D_P(z) = \inf_{u \in S^{d-1}} P\left(H(z,u)\right), \tag{1}$$

where $H(z,u) = \{x \in \mathbb{R}^d : \ x^T u \geq z^T u\}$ is the half-space passing through z in direction u, and S^{d-1} is the unit sphere in $\mathbb{R}^d$ with respect to the Euclidean norm $\|\cdot\|_2$. Let us remark that the notion of half-space depth and its applications appear in other areas, for example in graph theory (Small, 1997; Cerdeira and Silva, 2021), in convex geometry in relation to the convex floating boudies (Nagy et al., 2019), and the theory of social choice in economics under the name of the "min-max majority" (Caplin and Nalebuff, 1988; Nehring and Puppe, 2023). We refer the reader to the habilitation thesis of S. Nagy (2022) for an excellent overview of the history and recent developments in the theoretical and algorithmic aspects of Tukey's depth.

S. Minsker acknowledges support by the National Science Foundation grant DMS CAREER-2045068.

If P is such that $P(\partial H) = 0$ for the boundary ∂H of any closed half-space H, then the infumum in (1) is attained (Massé, 2004, Proposition 4.5). For example, this is the case for distributions that are absolutely continuous with respect to the Lebesgue measure. Of particular interest are the points of maximal depth, that is,

$$\mu_* := \mu_*(P) \in \arg\max D_P(z).$$

The set of deepest points with respect to P will be referred to as Tukey's median set, or simply Tukey's set, while the barycenter of this set is Tukey's median; when $d = 1$, it coincides with the standard median. According to Propositions 7 and 9 in (Rousseeuw and Ruts, 1999), Tukey's median always exists, and $\alpha_* := D_P(\mu_*) \geq \frac{1}{d+1}$ for any P. For distributions possessing symmetry properties, α_* can be much larger. For example, if P is halfspace-symmetric[1] and absolutely continuous with respect to the Lebesgue measure, then its center of symmetry coincides with Tukey's median, and its depth α_* equals $\frac{1}{2}$. The multivariate normal distribution, and more generally all elliptically symmetric distributions, are known to satisfy this property. Whenever α_* is large, Tukey's median also has a high breakdown point, and therefore features strong robustness properties – for instance, its breakdown point equals $\frac{1}{3}$ for halfspace-symmetric distributions, see Proposition 1 in (Chen, 1995) for the precise statement. The upper-level sets of half-space depth, defined as

$$R_P(\alpha) := \{z \in \mathbb{R}^d : \ D_P(z) \geq \alpha_* - \alpha\}, \ \alpha < \alpha_*,$$

are convex and compact. These sets are often called the *central regions* (Serfling, 2002) and their boundaries – the quantile surfaces or depth contours (Liu et al., 1999).

Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be a sequence of independent copies of a random vector X with distribution P. The empirical measure P_n^X is a discrete measure with atoms $X_1, \ldots, X_n$ having weight $1/n$ each. In particular, for any Borel measurable set $A \subseteq \mathbb{R}^d$,

$$P_n^X(A) := \frac{1}{n}\sum_{j=1}^{n} I_A(X_j), \ \text{where } I_A(z) = \begin{cases} 1, & z \in A, \\ 0, & z \notin A. \end{cases}$$

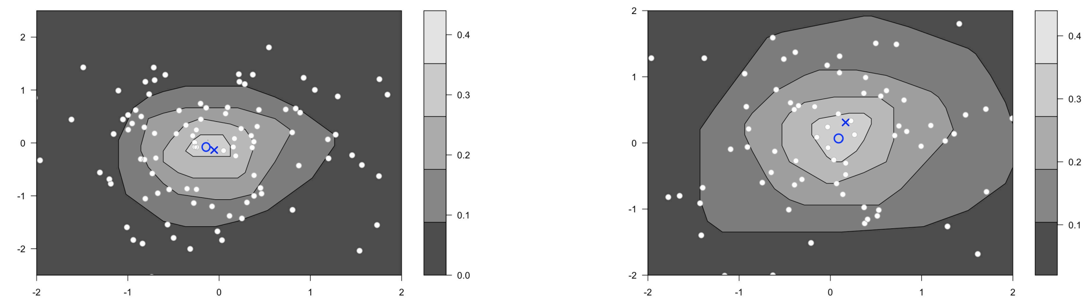
The *empirical depth* corresponding to P_n^X will be denoted $D_n(z)$, and it yields a natural way to order the points in a sample, giving rise to the depth-based ranks. It also admits the following convenient geometric interpretation: $nD_n(z)$ equals the cardinality of a smallest subset $S \subset \{X_1, \ldots, X_n\}$ such that z is not in the convex hull of $\{X_1, \ldots, X_n\} \setminus S$, with the convention that the cardinality of the empty set is 0. The empirical version of Tukey's set is defined as the set of points of maximal empirical depth,

$$\widehat{\mu}_n := \widehat{\mu}_n(X_1, \ldots, X_n) \in \arg\max_{z \in \mathbb{R}^d} D_n\left(z\right).$$

The barycenter of this set is known as the empirical Tukey's median, and it will also be denoted $\widehat{\mu}_n$ with slight abuse of notation. The definition is illustrated in Figure 1 that displays the (empirical) depth contours, upper-level sets and the Tukey's median corresponding to a sample of size $n = 100$ from the isotropic bivariate Gaussian distribution as well as the isotropic bivariate Student's t distribution with 2.1 degrees freedom. Figure 2 shows that the empirical contours become smoother and approach their population counterparts as the sample size grows.

Even when μ_* is unique, the Tukey's set can have a non-empty interior. For instance, if $d = 3$, then the set $\arg\max_{z \in \mathbb{R}^d} D_n\left(z\right)$ can be a single point, an interval, or a convex polytope, see example 3.1 in (Pokorný et al., 2024). This raises a natural question: *what is the typical diameter of this set?* In the univariate case ($d = 1$), the sample median is unique when n is odd, and when n is even, the diameter equals $L_n = \left|X_{\left(\frac{n}{2}+1\right)} - X_{\left(\frac{n}{2}\right)}\right|$. Under mild assumptions

[1]The distribution of a random vector X is halfspace-symmetric with respect to some $\mu \in \mathbb{R}^d$ if the random variables $u^T(X - \mu)$ and $u^T(\mu - X)$ are equidistributed for all $u \in S^{d-1}$.

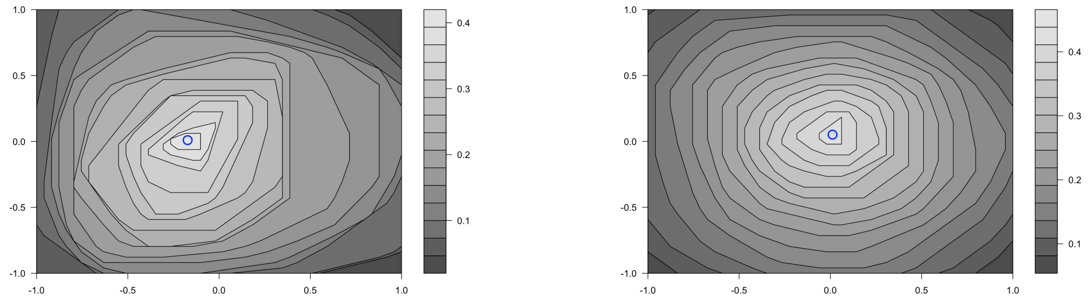


(A) $n = 100$, standard Gaussian

(B) $n = 100$, Student's t with $\nu = 2.1$ d.f.

FIGURE 1. Depth contours: lighter-colored regions correspond to higher depth. Empirical Tukey's median is marked with a circle and the sample mean – with a cross

.



(A) $n = 50$, standard Gaussian

(B) $n = 500$, standard Gaussian

FIGURE 2. Empirical depth contours approach their population limit. The blue circle denotes Tukey's median.

and using the classical results on order statistics (Rényi, 1953), the expected value of L_n can be shown to be of order c/n for a constant c that depends on the distribution. To the best of our knowledge, the question was open for $d \geq 2$. We make a step towards answering it and prove (see Corollary 2.3) that with high probability, diameter of empirical Tukey's set decays faster than $n^{-1/2}$. When $d = 2$, we obtain sharper estimates and show that the diameter is of order $n^{-3/4} \log^{3/2}$. These results follow from the general upper (Theorem 2.2) and lower (Theorem 3.1) bounds for the diameter of the sets of "deep points" that are proven to sharp in some cases.

Among the attractive properties of the half-space depth is the *affine invariance*: for any affine map $z \mapsto T(z) = Mz + b$ such that $M \in \mathbb{R}^{d\times d}$ is non-singular and $b \in \mathbb{R}^d$,

$$D_{P\circ T^{-1}}(Mz + b) = D_P(z), \ \forall z \in \mathbb{R}^d,$$

where $P \circ T^{-1}(A) = P\left(T^{-1}(A)\right)$. This implies that

$$\mu_*(P \circ T^{-1}) = M\mu_*(P) + b,$$

so that Tukey's median "respects" affine transformations. For example, if X has mean μ and the covariance matrix $\Sigma = \mathbb{E}(X-\mu)(X-\mu)^T$ is non-degenerate, then

$$\widehat{\mu}_n(X_1,\ldots,X_n) = \Sigma^{1/2}\widehat{\mu}_n(Z_1,\ldots,Z_n) + \mu,$$

where $Z_j = \Sigma^{-1/2}(X_j - \mu)$ has centered isotropic distribution (i.e. the covariance of Z_1 is the identity matrix). Due to this property, existing literature that investigates the proximity between the population median μ and its data-dependent counterpart $\widehat{\mu}_n$ often focuses on the estimation error measured in the Mahalanobis distance $\|\Sigma^{-1/2}(\widehat{\mu}_n(X_1,\ldots,X_n) - \mu)\|_2$. Since Σ is typically unknown, one may instead prefer to measure the error with respect to the usual Euclidean distance $\|\widehat{\mu}_n - \mu\|_2$.

In this paper, we establish certain optimality properties of Tukey's median. Specifically, we prove (see Theorem 2.1) that the size of the error $\|\widehat{\mu}_n - \mu\|_2$ is controlled by the *effective rank* of Σ defined as the ratio of the trace $\mathrm{tr}(\Sigma)$ and the spectral norm $\|\Sigma\|$,

$$r(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|},$$

as opposed to the ambient dimension d. Finally, we investigate robustness properties of Tukey's median and the effects of adversarial contamination on the error. While this question has been answered for isotropic distributions (equivalently, for the error with respect to the Mahalanobis norm, see (Chen et al., 2018)), the case of Euclidean distance poses new challenges. We give a partial answer in Corollary 2.4. The remaining open problems are discussed in the section 4.

1.1. **Notation.** Given $z_1, z_2 \in \mathbb{R}^d$, $\langle z_1, z_2\rangle$ will stand for the standard inner product, $\|\cdot\|_2$ – for the associated Euclidean norm and $B(z,r)$ will denote the Euclidean ball of radius r centered at z. For a matrix $A \in \mathbb{R}^{d\times d}$, $\mathrm{tr}(A)$ represents its trace and $\|A\|$ - its spectral (operator) norm. I_d will denote the $d\times d$ identity matrix.

We will employ the standard small-o and big-O asymptotic notation throughout the paper. Given two sequences $\{a_j\}_{j\geq 1},\ \{b_j\}_{j\geq 1} \subset \mathbb{R}_+$, we will say that $a_j \ll b_j$ if $a_j = o(b_j)$ as $j\to\infty$. We will write $o_P(1)$ to denote a sequence of random variables $\xi_j,\ j\geq 1$ that converge to 0 in probability as $j\to\infty$. Indicator function of an event $\mathcal{E}$ will be denoted $I_{\mathcal{E}}$ or $I\{\mathcal{E}\}$. Finally, $\Phi(t)$ will stand for the cumulative distribution function of the standard normal law on $\mathbb{R}$ and $\phi(t)$ – for the associated probability density function. Additional notation will be introduced on demand.

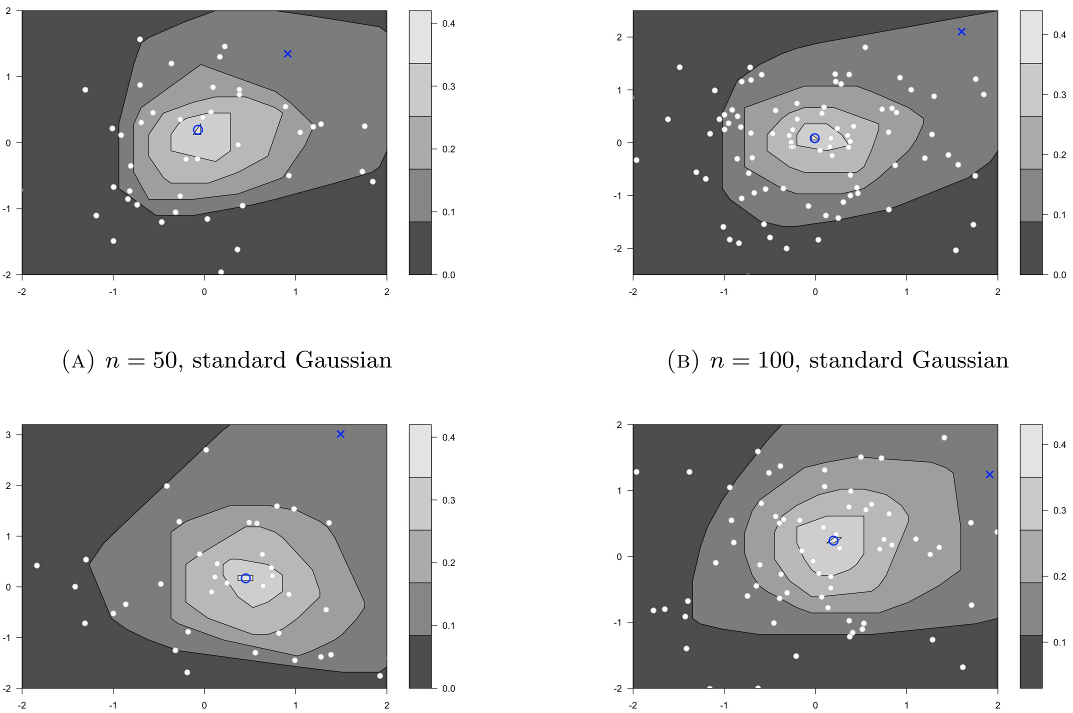
2. Main results

Assume that a fraction of the data has been replaced by arbitrary values, that is, we observe $Y_1,\ldots,Y_n$ where $\frac{1}{n}\sum_{j=1}^n I\{Y_j \neq X_j\} = \varepsilon$.[2] In a scenario when $X_1,\ldots,X_n$ have multivariate normal distribution $N(\mu,\Sigma)$ such that Σ is non-singular, Chen et al. (2018) showed that whenever $\varepsilon < 1/5$ and $\frac{d}{n} + \frac{t}{n} \leq c$,

$$\|\Sigma^{-1/2}\left(\widehat{\mu}_n(Y_1,\ldots,Y_n) - \mu\right)\|_2 \leq C\left(\sqrt{\frac{d+t}{n}} + \varepsilon\right), \tag{2}$$

holds with probability at least $1-e^{-t}$, where $c, C > 0$ are absolute constants. The key feature of this inequality is the fact that it implies optimal dependence of the error on the contamination proportion ε; Tukey's median is among the first estimators known to possess this property. This robustness property of Tukey's median is illustrated in figure 3 where it is juxtaposed with the sample mean.

[2]This is commonly referred to as the *adversarial contamination framework*, see (Diakonikolas and Kane, 2023)



(A) $n = 50$, standard Gaussian

(B) $n = 100$, standard Gaussian

(C) $n = 50$, Student's t with $\nu = 2.1$ d.f.

(D) $n = 100$, Student t with $\nu = 2.1$ d.f.

FIGURE 3. Contaminated sample with contamination proportion $\varepsilon = 0.1$. Circle mark denotes Tukey median and the cross represents the mean.

Recently, Minasyan and Zhivotovskiy (2023) proved that there exists an estimator $\tilde{\mu}_n$ that satisfies the inequality

$$\|\tilde{\mu}_n(Y_1, \ldots, Y_n) - \mu\|_2 \leq C\|\Sigma\|^{1/2}\left(\sqrt{\frac{r(\Sigma) + t}{n}} + \varepsilon\right) \tag{3}$$

with probability at least $1 - e^{-t}$, whenever $\frac{r(\Sigma)+t}{n} \leq c$. Furthermore, if $\|\Sigma\|$ and $r(\Sigma)$ are known up to some absolute multiplicative factor, then the requirement $\frac{r(\Sigma)+t}{n} \leq c$ can be dropped. This inequality exhibits optimal dependence on the "dimensional" characteristic expressed via the effective rank as opposed to the ambient dimension that can be much larger, as well as correct dependence on the degree of contamination ε. Deviation inequalities of type (3) are often referred to as *sub-Gaussian guarantees*, since they hold, for $\varepsilon = 0$, for the sample mean of i.i.d. Gaussian random vectors due to the Borell-TIS inequality (Borell, 1975; Tsirelson et al., 1976). Whether Tukey's median admits a bound of the form (3) remained an open question: specifically, it was unknown whether d can be replaced by $r(\Sigma)$. We will give a (partial) affirmative answer below. As this question is mainly interesting in the situation when d is large, we will assume that $d \geq 2$ in the remainder of the paper.

2.1. **Contamination-free framework.** Our starting point is the contamination-free scenario corresponding to $\varepsilon = 0$. Recall that a random vector $X \in \mathbb{R}^d$ has elliptically symmetric distribution $\mathcal{E}(\mu, \Sigma, F)$, denoted $X \sim \mathcal{E}(\mu, \Sigma, F)$ in the sequel, if

$$X \overset{d}{=} \eta \cdot BU + \mu,$$

where $\overset{d}{=}$ denotes equality in distribution, η is a nonnegative random variable with cumulative distribution function F, B is a fixed $d \times d$ matrix such that $\Sigma = BB^T$, and $U \in \mathbb{R}^d$ is uniformly distributed over the unit sphere $\mathbb{S}^{d-1}$ and independent of η. The distribution $\mathcal{E}(\mu,\ \mathbf{\Sigma},\ F)$ is well defined, as if $B_1 B_1^T = B_2 B_2^T$, then there exists an orthogonal matrix Q such that $B_1 = B_2 Q$, and $QU \overset{d}{=} U$. To avoid such ambiguity, we will allow B to be any matrix satisfying $BB^T = \Sigma$. Whenever $\mathbb{E}\eta^2 < \infty$, we will also assume that η is scaled so that $\mathbb{E}\eta^2 = 1$, in which case the covariance of X exists and is equal to Σ.

A useful property of elliptically symmetric distributions that is evident from the definition is the following: for every orthogonal matrix Q, $Q\Sigma^{-1/2}(X - \mu)$ has the same distribution as $\Sigma^{-1/2}(X-\mu)$. In other words, $\Sigma^{-1/2}(X-\mu)$ is spherically symmetric. We are ready to state our first result that essentially hinges on this fact.

Theorem 2.1. *Let $X_1, \ldots, X_n$ be i.i.d. copies of $X \sim \mathcal{E}(\mu, \Sigma, F)$ where Σ is non-singular. Let $\widetilde{\mu}_n$ be any affine equivariant estimator of μ. Suppose that*

$$\|\Sigma^{-1/2}(\widetilde{\mu}_n - \mu)\|_2 \le e(n,t,d)$$

holds with probability at least $1 - p(t)$. Then

$$\|\widetilde{\mu}_n - \mu\|_2 \le e(n,t,d)\|\Sigma\|^{1/2}\left(\sqrt{\frac{r(\Sigma)}{d}} + \sqrt{\frac{2t}{d}}\right)$$

with probability at least $1 - p(t) - e^{-t}$.

In typical scenarios,

$$e(n,t,d) \asymp \sqrt{\frac{d}{n}} + \sqrt{\frac{t}{n}} \text{ and } p(t) = e^{-t},$$

whence

$$\|\widetilde{\mu}_n - \mu\|_2 \le C\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{t}{n}}\right)\sqrt{\|\Sigma\|}\left(\sqrt{\frac{r(\Sigma)}{d}} + \sqrt{\frac{t}{d}}\right) \le C_1\sqrt{\|\Sigma\|}\left(\sqrt{\frac{r(\Sigma)}{n}} + \sqrt{\frac{t}{n}}\right) \tag{4}$$

with probability at least $1 - 2e^{-t}$ whenever $t \le d$. The proof of the theorem is given in appendix A. Next, we state the corollaries for Tukey's median and the Stahel-Donoho estimator, another well known affine-equivariant robust estimator of location (see (Stahel, 1981; Donoho, 1982) for its definition and properties).

Corollary 2.1 (Tukey's median)**.** *Let $X_1, \ldots, X_n$ be i.i.d. copies of $X \sim N(\mu, \Sigma)$ where Σ is non-singular, and let $\widehat{\mu}_n$ be the Tukey's median, the barycenter of the set $\arg\max_{z \in \mathbb{R}^d} D_n(z)$. Then there exist absolute constants $c, C > 0$ such that*

$$\|\widehat{\mu}_n - \mu\|_2 \le C\left(\sqrt{\frac{\operatorname{tr}(\Sigma)}{n}} + \sqrt{\|\Sigma\|}\sqrt{\frac{t}{n}}\right)$$

with probability at least $1 - 2e^{-t}$ whenever $\frac{d}{n} + \frac{t}{n} \le c$.

Proof. Theorem 2.1 in (Chen et al., 2018) states that

$$e(n,t,d) = \sqrt{\|\Sigma\|}\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{t}{n}}\right),$$

and $p(t) = e^{-t}$ whenever $\frac{d}{n} + \frac{t}{n} \le c$. Inequality (4) thus implies the desired result whenever $t \le d$. If $t > d$, it suffices to show that under the stated assumptions, $\|\Sigma^{-1/2}(\widehat{\mu}_n - \mu)\| \le C\sqrt{\frac{t}{n}}$ with probability at least $1 - 2e^{-t}$, which follows directly from Theorem 2.1 in (Chen et al., 2018). □

Remark 2.1. *The bound of Corollary 2.1 remains valid whenever* $d \ge n+1$*: indeed, in this case with probability* 1 *the convex hull of* $X_1, \ldots, X_n$ *is a simplex with* n *vertices, and every point in the convex hull has (maximal) depth* $1/n$*, hence* $\widehat{\mu}_n$ *coincides with the sample mean.*

Up to the values of numerical constants, the same result is valid for the Stahel-Donoho estimator.

Corollary 2.2 (Stahel-Donoho estimator)**.** *Let* $X_1, \ldots, X_n$ *be i.i.d. copies of* $X \sim N(\mu, \Sigma)$ *where* Σ *is non-singular, and let* $\widetilde{\mu}_n$ *be the barycenter of the set of Stahel-Donoho estimators. Then there exist absolute constants* $c, C > 0$ *such that*

$$\|\widetilde{\mu}_n - \mu\|_2 \le C\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{n}} + \sqrt{\|\Sigma\|}\sqrt{\frac{t}{n}}\right)$$

with probability at least $1 - 2e^{-t}$ *whenever* $\frac{d}{n} + \frac{t}{n} \le c$.

Proof. The argument repeats the proof of Corollary 2.1 *mutatis mutandis*, where Theorem 1 in (Depersin and Lecué, 2023) plays the role of Theorem 2.1 in (Chen et al., 2018). □

Remark 2.2. *Results of both corollaries hold not just for Gaussian measures but more generally for a wide class of elliptically symmetric distributions. Indeed, the arguments in (Chen et al., 2018; Depersin and Lecué, 2023) leading to the bounds for* $e(n, t, d)$ *remain valid if we assume that* $X \sim \mathcal{E}(\mu, \Sigma, F)$ *is such that the density* $q(t)$ *of one-dimensional projections*

$$\left\langle \Sigma^{-1/2}(X - \mu), v \right\rangle = \eta \langle U, v \rangle,$$

where $\langle \cdot, \cdot \rangle$ *denotes the standard Euclidean inner product, is continuous and strictly positive at the origin (also see remark 2.2 in (Chen et al., 2018)). For example, multivariate t-distribution satisfies this assumption.*

2.2. **Performance guarantees in the adversarial contamination framework.** Recall the adversarial contamination framework that was introduced in the beginning of section 2. All results stated below will be specific to Tukey's median as we rely on the existing rich literature on the subject. However, we expect our claims to be valid for other affine equivariant estimators such as the Stahel-Donoho estimator mentioned previously.

In what follows, we will let $\widehat{\mu}_n^X$ be the empirical Tukey's median based on the uncontaminated data $X_1, \ldots, X_n$ and $\widehat{\mu}_n^Y$ – its counterpart based on the sample $Y_1, \ldots, Y_n$. We will let

$$D_n^X(z) = \inf_{\|u\|_2 = 1} P_n^X H(z, u)$$

stand for the empirical depth of a point $z \in \mathbb{R}^d$, where the empirical distribution P_n^X is based on $X_1, \ldots, X_n$, and $\widehat{d}_n := D_n^X\left(\widehat{\mu}_n^X\right)$ be the maximal (empirical) depth. Consider the set

$$R_n(\delta) := \{z \in \mathbb{R}^d : \ D_n^X(z) \ge \widehat{d}_n - \delta\}$$

of "$\widehat{d}_n - \delta$-deep" points. $R_n(\delta)$ is a convex polytope whose facets belong to the hyperplanes defined by the elements of the sample. Observe that the empirical depth $D_n^Y(z)$ (based on the contaminated dataset) of any point $z \notin R_n(\delta)$ satisfies the inequality

$$D_n^Y(z) < \varepsilon + \widehat{d}_n - \delta.$$

Indeed,

$$D_n^Y(z) = \inf_{\|u\|_2=1} P_n^Y H(z,u) \le \underbrace{\inf_{\|u\|_2=1} P_n^X H(z,u)}_{<\widehat{d}_n-\delta} + \underbrace{\sup_{\|u\|_2=1} \left|\left(P_n^Y - P_n^X\right) H(z,u)\right|}_{\le\varepsilon}. \tag{5}$$

The latter does not exceed $\widehat{d}_n - \varepsilon$ whenever $\delta > 2\varepsilon$. Since $D_n^Y(\widehat{\mu}_n)$ is no smaller than $\widehat{d}_n - \varepsilon$ due to the inequality in the spirit of (5), we deduce the following lemma.

Lemma 2.1. *Let $\widehat{\mu}_N^Y \in \arg\max_{z\in\mathbb{R}^d} D_n^Y(z)$. Then*

$$\widehat{\mu}_n^Y \in R_n(2\varepsilon).$$

An immediate corollary of this lemma is that in the framework of Corollary 2.1,

$$\|\widehat{\mu}_n^Y - \mu\|_2 \le \|\widehat{\mu}_n^X - \mu\|_2 + \mathrm{diam}\left(R_n(2\varepsilon)\right), \tag{6}$$

where $\mathrm{diam}(A)$ stands for the diameter of a set $A \subseteq \mathbb{R}^d$. Whenever ε is large, meaning that $1/5 > \varepsilon \ge K\sqrt{\frac{d}{n}}$ for some constant $K > 0$, results by Brunel (2019) imply that the Hausdorff distance between $R_n(\varepsilon)$ and its population analogue $R(\varepsilon)$ is of order $O_P\left(\sqrt{\|\Sigma\|}\sqrt{\frac{d}{n}}\right)$, hence it follows that

$$\mathrm{diam}\left(R_n(2\varepsilon)\right) \le C(K)\sqrt{\|\Sigma\|}\left(\sqrt{\frac{t}{n}} + \varepsilon\right)$$

with probability at least $1 - 2e^{-t}$ for $t \le c_1 n$ (alternatively, this inequality can be obtained from the bound (2) in the case of normally distributed data). Therefore, for such values of ε and t,

$$\left\|\widehat{\mu}_n^Y - \mu\right\|_2 \le C(K)\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{n}} + \sqrt{\|\Sigma\|}\left(\sqrt{\frac{t}{n}} + \varepsilon\right)\right) \tag{7}$$

with probability at least $1 - 2e^{-t}$.

The problem is less straightforward for $\varepsilon := \varepsilon_n \ll \sqrt{\frac{d}{n}}$. In the remainder of this section, we will assume that d and $r(\Sigma)$ are fixed, where d is potentially much larger than $r(\Sigma)$, and $n \to \infty$. The following theorem is the main result of this section.

Theorem 2.2. *Assume that $X_1, \dots, X_n$ are independent random vectors sampled from the isotropic normal distribution $N(\mu, I_d)$. Then*

(i) for any sequence $\varepsilon_n = o(n^{-1/2})$, the diameter of the set $R_n(\varepsilon_n)$ satisfies

$$\mathrm{diam}(R_n(\varepsilon_n)) = o_P\left(n^{-1/2}\right).$$

(ii) if $d = 2$, then

$$\mathrm{diam}(R_n(\varepsilon_n)) = O_P\left(n^{-\frac{3}{4}}\log^{3/2}(n) \vee \varepsilon_n\right).$$

In section 3, we will show that the bound **(ii)** is sharp for $\varepsilon_n \gtrsim n^{-\frac{3}{4}}\log^{3/2}(n)$. Before turning to the proof of this result, we will state its immediate corollaries.

Corollary 2.3 (Tukey's median set)**.** *The diameter of the set $R_n(0)$ satisfies*

$$\mathrm{diam}(R_n(0)) = o_P\left(n^{-1/2}\right).$$

Moreover, when $d = 2$,

$$\mathrm{diam}(R_n(0)) = O_P\left(n^{-\frac{3}{4}}\log^{3/2}(n)\right).$$

Corollary 2.4 (Adversarial Contamination)**.** *Assume that $X_1, \ldots, X_n$ are i.i.d. random vectors sampled from the multivariate normal distribution $N(\mu, \Sigma)$. In the adversarial contamination framework with $\varepsilon_n = o(n^{-1/2})$,*

$$\left\|\widehat{\mu}_n^Y - \mu\right\|_2 \le C\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{n}} + \sqrt{\|\Sigma\|}\sqrt{\frac{t}{n}}\right)$$

with probability at least $1 - 2e^{-t} - o(1)$, where $C > 0$ is an absolute constant.

Proof. First, note that due to affine invariance of the depth function $D_n(z)$, the depth regions corresponding to the data $X_1, \ldots, X_n)$ and $\Sigma^{-1/2}(X_1 - \mu), \ldots, \Sigma^{-1/2}(X_n - \mu)$ are related via

$$R_n(2\varepsilon_n; X_1, \ldots, X_n) = \mu + \Sigma^{1/2}\, R_n(2\varepsilon_n; \Sigma^{-1/2}(X_1 - \mu), \ldots, \Sigma^{-1/2}(X_n - \mu)).$$

Therefore,

$$\mathrm{diam}\left(R_n(2\varepsilon_n; X_1, \ldots, X_n)\right) \le \|\Sigma\|^{1/2}\mathrm{diam}\left(R_n(2\varepsilon_n; \Sigma^{-1/2}(X_1 - \mu), \ldots, \Sigma^{-1/2}(X_n - \mu))\right).$$

The latter diameter is $o_P(n^{-1/2})$ in view of Theorem 2.2. The inequality (6) combined with Theorem 2.2 and Corollary 2.1 yields that

$$\left\|\widehat{\mu}_n^Y - \mu\right\|_2 \le C\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{n}} + \sqrt{\|\Sigma\|}\sqrt{\frac{t}{n}}\right) + o_P(n^{-1/2}) \tag{8}$$

with probability at least $1 - 4e^{-t} - \frac{2}{n}$. For sufficiently large n, the summand $o_P(n^{-1/2})$ in (8) is dominated by the first term, leading to the desired conclusion. □

Of course, for $\varepsilon_n = o(n^{-1/2})$, the corollary trivially yields that

$$\left\|\widehat{\mu}_n^Y - \mu\right\|_2 \le C\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{n}} + \sqrt{\|\Sigma\|}\left(\sqrt{\frac{t}{n}} + \varepsilon_n\right)\right)$$

with high probability. In view of (7), this inequality is also valid for large values of ε_n, above $K\sqrt{\frac{d}{n}}$. Whether a bound of this form also holds whenever $\varepsilon_n \asymp \sqrt{\frac{r(\Sigma)}{n}}$ (or more generally in the regime $C_1\sqrt{\frac{r(\Sigma)}{n}} \le \varepsilon_n \le C_2\sqrt{\frac{d}{n}}$, where $C_1, C_2 > 0$ are some fixed constants), remains an open problem. Next, we turn our attention to the proof of Theorem 2.2.

Proof of Theorem 2.2. Everywhere below, we will assume without loss of generality that $\mu = 0$. We will first establish claim **(i)** of the Theorem. The idea of the proof relies on the strong approximation of the depth process via a P-Brownian bridge that allows us to relate the behavior of $W_n(z)$ and $W(z)$ near their respective points of maxima. We will need several preliminary facts. Let $C > 0$ be a sufficiently large absolute constants. Recall that in view of Theorem 2.1 in Chen et al. (2018), for $\varepsilon_n = o(n^{-1/2})$,

$$R_n(2\varepsilon_n) \subseteq B\left(0, C_1\sqrt{\frac{d+t}{n}}\right)$$

with probability at least $1 - e^{-t}$. Next, define

$$W_n(z) = \sqrt{n}\inf_{\|v\|_2 = 1}\left(P_n^X H\left(\frac{z}{\sqrt{n}}, v\right) - \frac{1}{2}\right),\ \|z\|_2 \le C_1\sqrt{d+t}. \tag{9}$$

On event of probability at least $1 - e^{-t}$, we can equivalently express $R_n(2\varepsilon_n)$ via

$$R_n(2\varepsilon_n) = \left\{\frac{z}{\sqrt{n}} :\ \|z\|_2 \le C\sqrt{d+t} \text{ and } W_n(z) \ge \sup_{z\in\mathbb{R}^d} W_n(z) - 2\sqrt{n}\varepsilon_n\right\}.$$

Moreover, the depth process $W_n(z)$ converges weakly to

$$W(z) := W_G(z) = \inf_{\|v\|_2=1} \left\{ G(v,0) - \frac{1}{\sqrt{2\pi}}\langle z, v\rangle \right\}, \ \|z\|_2 \le C_1\sqrt{d+t}, \tag{10}$$

where $G(v, y)$ is the Brownian bridge indexed by half-spaces – a centered Gaussian process with covariance function

$$\mathbb{E}G(u,z)G(v,y) = \mathrm{Cov}\left(I\{X \in H(z,u)\}, I\{X \in H(y,v)\}\right),$$

where $X \sim N(0, I_d)$. This result is known, for example see (Massé, 2002, page 296). We will also need the following estimate on the rate of convergence of W_n to W.

Lemma 2.2. *For every $C > 0$, there exists $C_1 > 0$ such that for all $n > C_1 R$, one can construct on the same probability space independent random vectors $X_1, \ldots, X_n$ with standard normal distribution and a Brownian bridge G such that*

$$\sup_{\|z\|_2 \le R} |W_n(z) - W(z)| \le \delta_n + K\left(\frac{R^3}{n} + \left(\frac{R}{n}\right)^{1/4}\sqrt{\log(n)}\right)$$

with probability at least $1-2/n$, where $K > 0$ is an absolute constant and $\delta_n = C(d)n^{-\frac{1}{2d}}\log^{3/2}(n)$.

We include the proof of the lemma in the appendix. In what follows, we will set $R = d + t$ and

$$\gamma_n(t) := \delta_n + K\left(\frac{(d+t)^3}{n} + \left(\frac{d+t}{n}\right)^{1/4}\sqrt{\log(n)}\right).$$

The final ingredient of the proof is the following statement.

Lemma 2.3. *Let $R(\beta) = \{\|z\|_2 \in \mathbb{R}^d : \ W(z) \ge \sup_{z\in\mathbb{R}^d} W(z) - \beta\}$. Then there exists event $\mathcal{E}$ of probability at least $1 - e^{-t}$ such that for any sequence $\beta_j = o(1)$ and any $\tau > 0$,*

$$\lim_{j\to\infty} \mathbb{P}\left(\{\mathrm{diam}(R(\beta_j)) \ge \tau\} \cap \mathcal{E}\right) = 0. \tag{11}$$

Proof. Lemma B.1 (i) guarantees that $\widehat{z}$ is unique and that $\|\widehat{z}\|_2 \le C\sqrt{d+t}$ on event $\mathcal{E}$ of probability at least $1 - e^{-t}$. Assume that (11) is false, meaning that for some increasing sequence $\{j_i\}_{i\ge1}$,

$$\lim_{i\ge1} \mathbb{P}\left(\{\mathrm{diam}(R(\beta_{j_i})) \ge \tau\} \cap \mathcal{E}\right) \ge c > 0.$$

Since the sets $R(\beta_{j_i})$ are nested, it implies that

$$\mathbb{P}\left(\left\{\mathrm{diam}\left(\bigcap_{i\ge1} R\left(\beta_{j_i}\right)\right) \ge \tau\right\} \cap \mathcal{E}\right) > 0.$$

But any element of the set $\bigcap_{i\ge1} R(\beta_{j_i})$ must be the $\arg\max_z W(z)$, contradicting the a.s. uniqueness of the latter. The claim follows. □

We will now deduce the statement of the theorem. It follows from the facts established above that there exists an event Θ of probability at least $1 - 2e^{-t} - \frac{2}{n}$ such that the claims of Lemmas 2.2 and 2.3 hold on Θ, and moreover the condition $R_n(2\varepsilon_n) \subseteq B\left(0, C_1\sqrt{\frac{d+t}{n}}\right)$ is satisfied.

Therefore, on Θ

$$
\begin{aligned}
(12)\quad R_n(\varepsilon_n) &= \left\{ \frac{z}{\sqrt{n}} : \|z\|_2 \le C_1\sqrt{d+t} \text{ and } W_n(z) \ge \sup_{z\in\mathbb{R}^d} W_n(z) - 2\sqrt{n}\varepsilon_n \right\} \\
&\subseteq \left\{ \frac{z}{\sqrt{n}} : \|z\|_2 \le C_1\sqrt{d+t},\, W(z) \ge \sup_{z\in\mathbb{R}^d} W(z) - 2\sup_{\|z\|\le C_1\sqrt{d+t}} |W_n(z) - W(z)| - 2\sqrt{n}\varepsilon_n \right\} \\
&\subseteq \left\{ \frac{z}{\sqrt{n}} : \|z\|_2 \le C_1\sqrt{d+t} \text{ and } W(z) \ge \sup_{z\in\mathbb{R}^d} W(z) - 2\gamma_n(t) - 2\sqrt{n}\varepsilon_n \right\} := \mathcal{A}_n.
\end{aligned}
$$

As $\gamma_n(t) + \sqrt{n}\varepsilon_n = o(1)$ as $n \to \infty$, we conclude that

$$\sqrt{n}\,\mathrm{diam}(\mathcal{A}_n) \cdot I_\Theta \to 0 \text{ in probability as } n \to \infty.$$

Since t can be arbitrarily large, the claim follows.

To prove part **(ii)** of the theorem, we use similar reasoning combined with claim **(ii)** of Lemma B.1. The latter states that with probability 1, $\sup_{w\neq 0} \min_{u\in\mathcal{U}} u^\top \frac{w}{\|w\|_2} < 0$ where $\mathcal{U}$ is the set of extreme points of $\partial W(\widehat{z})$. Therefore, for any $\alpha > 0$, there exists $c(\alpha) > 0$ such that

$$\min_{u\in\mathcal{U}} u^\top \frac{w}{\|w\|_2} \le -c(\alpha)$$

with probability at least $1-\alpha$. The trajectories of $W(z)$ are concave with probability 1 as the pointwise infimum of linear functions, therefore, for any $z \neq \widehat{z}$ and any $u \in \mathcal{U}$

$$W(z) \le W(\widehat{z}) + u^\top (z - \widehat{z}).$$

Choosing $u = u(z) \in \mathcal{U}$ for which $u^\top(z-\widehat{z})$ is minimized, we deduce that $W(z) \le W(\widehat{z}) - c(\alpha)\|z-\widehat{z}\|_2$ with probability at least $1-\alpha$. In view of (12), this yields that

$$
\begin{aligned}
R_n(\varepsilon_n) &\subseteq \left\{ \frac{z}{\sqrt{n}} :\; c(\alpha)\|z-\widehat{z}\|_2 \le 2\gamma_n(t) + 2\sqrt{n}\varepsilon_n \right\} \\
&\subseteq \left\{ z :\; \left\| z - \frac{\widehat{z}}{\sqrt{n}} \right\|_2 \le \frac{4}{c(\alpha)\sqrt{n}} \left(\gamma_n(t) \vee \sqrt{n}\varepsilon_n\right) \right\}
\end{aligned}
$$

on event of probability at least $1 - 2e^{-t} - \frac{2}{n} - \alpha$. Setting $d = 2$ in the definition of γ_n, we obtain the desired conclusion. □

3. Lower bounds for the diameter of the empirical depth regions

The goal of this section is to establish lower bounds on the diameter of the sets $R_n(\varepsilon)$, and to show that the upper bounds obtained before for the case $d = 2$ are optimal in some regimes.

Theorem 3.1. *With probability at least* $1 - e^{-t}$,

$$
\begin{aligned}
\sup_{\|z_1 - z_2\|_2 \le \varepsilon} |D_n(z_1) - D_n(z_2)| &\le \frac{\varepsilon}{\sqrt{2\pi}} \\
&+ C_1 \left(\sqrt{\varepsilon} \left(\sqrt{\frac{t}{n}} + \sqrt{\frac{d}{n}} \log^{1/2}\left(\frac{C_2}{\varepsilon}\right) \right) + \frac{1}{n}\left(t + d\log\left(\frac{C_2}{\varepsilon}\right)\right)\right).
\end{aligned}
$$

The following corollary states that the depth regions are not too small: in particular, diameter of $R_n(\varepsilon)$ grows at least linearly with respect to ε when ε. Given two subsets $A, B \subseteq R^d$, $A + B = \{a + b,\ a \in A,\ b \in B\}$ stands for their Minkowski sum.

Corollary 3.1. *There exist absolute constants c, c_1, C_3 with the following property. Assume that $\frac{d+t}{n} \le c$ for c small enough. Then for any $\varepsilon \ge C_3 \frac{d+t}{n} \log\left(\frac{n}{d+t}\right)$,*

$$R_n(\varepsilon) \supseteq R_n(0) + B(0, c_1\varepsilon)$$

with probability at least $1 - e^{-t}$, where $R_n(0)$ is the set of all Tukey's medians.

In particular, we immediately see for $d = 2$,

$$\mathrm{diam}(R_n(\varepsilon)) \ge 2c_1\varepsilon$$

for ε larger than $\frac{C}{n}\log(n)$. On the other hand, result of Theorem 2.2 part **(ii)** states that with high probability,

$$\mathrm{diam}(R_n(\varepsilon)) \le C'\varepsilon$$

whenever $\varepsilon \ge C'' n^{-3/4} \log^{3/2}(n)$. Therefore, for ε that exceed $C'' n^{-3/4} \log^{3/2}(n)$, the bound of Theorem 2.2 part **(ii)** is sharp.

Proof of Corollary 3.1. It is easy to see that whenever $\varepsilon \ge C_3 \frac{d+t}{n} \log\left(\frac{n}{d+t}\right)$, the term $\frac{\varepsilon}{\sqrt{2\pi}}$ dominates in the bound of Theorem 3.1. Therefore, $D_n(z) \ge D_n(\widehat{\mu}_n) - \varepsilon$ whenever $\|z - \widehat{\mu}_n\|_2 \le c_1\varepsilon$ for a sufficiently small constant $c_1 > 0$. □

Proof of Theorem 3.1. Given $z_1, z_2 \in \mathbb{R}^d$ such that $\|z_1 - z_2\|_2 \le \varepsilon$, let v_1, v_2 be unit vectors such that $D_n(z_j) = P_n^X H(z_j, v_j)$, $j = 1, 2$. Assuming without loss of generality that $D_n(z_1) \ge D_n(z_2)$, note that

$$\begin{aligned} D_n(z_1) - D_n(z_2) &= \underbrace{P_n^X \left(H(z_1, v_1) - H(z_1, v_2)\right)}_{\le 0} + P_n\left(H(z_1, v_2) - H(z_2, v_2)\right) \\ &\le \sup_{\|v\|_2=1, \|z_1 - z_2\| \le \varepsilon} \left|(P_n - P)(H(z_1, v) - H(z_2, v))\right| \\ &\qquad + \sup_{\|v\|_2=1, \|z_1 - z_2\| \le \varepsilon} \left|P(H(z_1, v) - H(z_2, v))\right|. \end{aligned} \tag{13}$$

We will estimate both terms on the right side of the display above. Denote

$$F_n(z, v) = \sqrt{n}(P_n^Z - P) H(z, v).$$

The following lemma controls its modulus of continuity.

Lemma 3.1. *There exist absolute constants $C_1, C_2 > 0$ such that for any $\varepsilon > 0$,*

$$\begin{aligned} &\sup_{\|z_1 - z_2\|_2 \le \varepsilon, \|v\|_2 = 1} |F_n(z_1, v) - F_n(z_2, v)| \\ &\qquad \le C_1 \left(\sqrt{\varepsilon}\left(\sqrt{t} + \sqrt{d}\log^{1/2}\left(\frac{C_2}{\varepsilon}\right)\right) + \frac{1}{\sqrt{n}}\left(t + d\log\left(\frac{C_2}{\varepsilon}\right)\right)\right) \end{aligned}$$

with probability at least $1 - e^{-t}$.

Proof. For brevity, set

$$S := \sup_{\|z_1 - z_2\|_2 \le \varepsilon, \|v\|_2 = 1} |F_n(z_1, v) - F_n(z_2, v)|,$$

where the supremum is taken over all z_1, z_2 such that $\|z_1 - z_2\|_2 \le \varepsilon$ and all unit vectors v. Note that

$$
\begin{aligned}
(14) \qquad & \sup_{\|z_1-z_2\|_2\le\varepsilon, \|v\|_2=1} P(H(z_1,v) - H(z_2,v))^2 \\
& = \sup_{\|z_1-z_2\|_2\le\varepsilon} \sup_{\|v\|_2=1} \left| \int_{\langle z_2, v\rangle}^{\langle z_1, v\rangle} \phi(z) dz \right| \le \sup_{\|z_1-z_2\|_2\le\varepsilon} \frac{\|z_1-z_2\|_2}{\sqrt{2\pi}} = \frac{\varepsilon}{\sqrt{2\pi}}.
\end{aligned}
$$

where $\phi(z)$ is the probability density function of the standard normal law. Bousquet's form of Talagrand's concentration inequality (Bousquet, 2003, Theorem 7.3) yields that

$$
S \le C\left(\mathbb{E}S + \sqrt{\varepsilon}\sqrt{t} + \frac{t}{\sqrt{n}} \right)
$$

with probability at least $1 - e^{-t}$. It remains to estimate for $\mathbb{E}S$. Since the class of half-spaces has Vapnik-Chervonenkis dimension $d+1$ (van der Vaart and Wellner, 2023), a well known argument (for instance, see Theorem 3.12 in Koltchinskii (2011)) implies that

$$
\mathbb{E}S \le C_1 \left(\sqrt{d\varepsilon} \log^{1/2}\left(\frac{C_2}{\varepsilon}\right) + \frac{d}{\sqrt{n}} \log\left(\frac{C_2}{\varepsilon}\right) \right)
$$

for some absolute constants $C_1, C_2 > 0$. □

Next, we deduce in a way similar to (14) that

$$
\sup_{\|v\|_2=1, \|z_1-z_2\|\le\varepsilon} |P(H(z_1,v) - H(z_2,v))| \le \frac{\varepsilon}{\sqrt{2\pi}}.
$$

Therefore, (13) implies that with probability at least $1 - e^{-t}$,

$$
\begin{aligned}
\sup_{\|z_1-z_2\|_2\le\varepsilon} |D_n(z_1) - D_n(z_2)| \le & \frac{\varepsilon}{\sqrt{2\pi}} \\
& + C_1 \left(\sqrt{\varepsilon}\left(\sqrt{\frac{t}{n}} + \sqrt{\frac{d}{n}} \log^{1/2}\left(\frac{C_2}{\varepsilon}\right) \right) + \frac{1}{n}\left(t + d\log\left(\frac{C_2}{\varepsilon}\right) \right) \right)
\end{aligned}
$$

which is the desired bound. □

4. Discussion

We proved that Tukey's median, and more generally affine equivariant estimators of location, are sensitive to the geometry of the underlying distribution, in a sense that their performance, measured with respect to the Euclidean distance, is controlled by the effective rank of the "shape" matrix rather than the ambient dimension. In the adversarial contamination framework, our claim remains valid as long as the contamination proportion ε is either large (larger than $\sqrt{\frac{d}{n}}$) or small, namely of order $o(n^{-1/2})$. There is still a significant gap between the bounds for Tukey's median proved in this paper and best known guarantees for the alternative estimators, for example by Minasyan and Zhivotovskiy (2023). This fact motivates the following questions:

(1) Is it possible to prove a version of Corollary 2.4 that allows the effective rank $r(\Sigma)$ and the ambient dimension d grow with the sample size n?
(2) What are the best guarantees for Tukey's median in the adversarial contamination framework with ε_n of order $\sqrt{\frac{r(\Sigma)}{n}}$?
(3) More generally, does Tukey's median satisfy non-asymptotic, finite-sample guarantees of the form (3)?

(4) Can the bound of Corollary 2.3 be sharpened? In other words, what is the typical diameter of Tukey's median set when the data follow an isotropic normal distribution? Formally following the steps of our analysis in the case $d = 1$, it is possible to obtain the rate for the diameter of the set of medians $R_n(0)$ of order $\frac{\log(n)}{n}$ which is sharp up to log-factors. We conjecture that in the case $d = 2$, our result is also nearly sharp, namely that $\mathrm{diam}(R_n(0))$ is of order $n^{-3/4}$ up to log-factors.

References

Aubrun, G., Jenkinson, J., and Szarek, S. J. (2024). Optimal constants in concentration inequalities on the sphere and in the Gauss space. *arXiv preprint arXiv:2406.13581*.

Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Inventiones mathematicae*, 30(2):207–216.

Bousquet, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, pages 213–247. Springer.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Brown, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 45(1):25–30.

Brunel, V.-E. (2019). Concentration of the empirical level sets of Tukey's halfspace depth. *Probability Theory and Related Fields*, 173(3):1165–1196.

Caplin, A. and Nalebuff, B. (1988). On 64%-majority rule. *Econometrica: Journal of the Econometric Society*, pages 787–814.

Cerdeira, J. O. and Silva, P. C. (2021). A centrality notion for graphs based on Tukey depth. *Applied Mathematics and Computation*, 409:126409.

Chen, M., Gao, C., and Ren, Z. (2018). Robust covariance and scatter matrix estimation under huber's contamination model. *The Annals of Statistics*, 46(5):1932–1960.

Chen, Z. (1995). Robustness of the half-space median. *Journal of statistical planning and inference*, 46(2):175–181.

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45(1):223–256.

Depersin, J. and Lecué, G. (2023). On the robustness to adversarial corruption and to heavy-tailed data of the Stahel–Donoho median of means. *Information and Inference: A Journal of the IMA*, 12(2):814–850.

Diakonikolas, I. and Kane, D. M. (2023). *Algorithmic high-dimensional robust statistics*. Cambridge university press.

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Harvard University. *Statistics Dept.*

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, pages 1803–1827.

Dudley, R. M. (1967). The sizes of compact subsets of hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis*, 1(3):290–330.

Dudley, R. M. (2014). *Uniform central limit theorems*, volume 142. Cambridge university press.

Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *The Annals of Statistics*, 30(4):1103–1133.

Hodges, J. L. (1955). A bivariate sign test. *The Annals of Mathematical Statistics*, 26(3):523–527.

Hotelling, H. (1929). Stability in competition. *The economic journal*, 39(153):41–57.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics; 2nd ed.* Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ.

Koltchinskii, V. (2011). *Oracle inequalities in empirical risk minimization and sparse recovery problems: École D'Été de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. Springer Science & Business Media.

Koltchinskii, V. I. (1994). Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. *Journal of Theoretical Probability*, 7:73–118.

Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, pages 435–477.

Lévy, P. (1951). Problèmes concrets d'analyse fonctionnelle. *2nd ed. Gauthier-Villars, Paris.*

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858.

Massé, J.-C. (2002). Asymptotics for the Tukey median. *Journal of Multivariate Analysis*, 81(2):286–300.

Massé, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10(3):397–419.

Minasyan, A. and Zhivotovskiy, N. (2023). Statistically optimal robust mean and covariance estimation for anisotropic Gaussians. *arXiv preprint arXiv:2301.09024*.

Nagy, S. (2022). Halfspace depth: Theory and computation. *Habilitation Thesis.*

Nagy, S., Schuett, C., and Werner, E. (2019). Data depth and floating body. *Statistics surveys*, 13.

Nehring, K. and Puppe, C. (2023). Multi-dimensional social choice under frugal information: the Tukey median as Condorcet winner ex ante by. Technical report, KIT Working Paper Series in Economics.

Nolan, D. (1999). On min–max majority and deepest points. *Statistics & probability letters*, 43(4):325–333.

Pokornỳ, D., Laketa, P., and Nagy, S. (2024). Another look at halfspace depth: flag halfspaces with applications. *Journal of Nonparametric Statistics*, 36(1):165–181.

Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung*, 4(2):48–89.

Rousseeuw, P. J. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49(3):213–244.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232.

Small, C. G. (1997). Multidimensional medians arising from geodesics on graphs. *The Annals of Statistics*, pages 478–494.

Stahel, W. A. (1981). *Breakdown of covariance estimators.* Fachgruppe für Statistik, Eidgenössische Techn. Hochsch.

Tsirelson, B., Ibragimov, I., and Sudakov, V. (1976). Norms of Gaussian sample functions. In *Proceedings of the Third Japan—USSR Symposium on Probability Theory*, pages 20–41. Springer.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, pages 523–531.

van der Vaart, A. and Wellner, J. (2023). *Weak convergence and empirical processes: with applications to statistics.* Springer Science & Business Media.

Appendix A. Proof of Theorem 2.1.

Define the random vector

$$W := \Sigma^{-1/2}(\widetilde{\mu}_n(X_1, \ldots, X_n) - \mu),$$

and observe that we need to estimate $\|\Sigma^{1/2}W\|_2$. Moreover, set $Z_j = \Sigma^{-1/2}(X_j-\mu),\ j=1,\dots,n$. Let us record the following identity:

$$\Sigma^{1/2}W = \|W\|_2\Sigma^{1/2}\left[\frac{W}{\|W\|_2}\right],$$

where $U := \frac{W}{\|W\|_2}$ is defined as 0 if $W=0$. We claim that conditionally on $W \neq 0$, U has uniform distribution over the sphere of radius 1. Indeed, for any orthogonal matrix Q, the vector QZ_1 has the same distribution as Z_1 in view of elliptical symmetry, which implies that $\widetilde{\mu}_n(QZ_1,\dots,QZ_n)$ and $\widetilde{\mu}_n(Z_1,\dots,Z_n)$ have the same distribution. But

$$\widetilde{\mu}_n(QZ_1,\dots,QZ_n) = Q\widetilde{\mu}_n(Z_1,\dots,Z_n),$$
$$\widetilde{\mu}_n(Z_1,\dots,Z_n) = \Sigma^{-1/2}(\widetilde{\mu}_n(X_1,\dots,X_n)-\mu) = W,$$

due to affine equivariance, hence $W \overset{d}{=} QW$. It implies that the conditional distribution of $\frac{W}{\|W\|}$ given that $W \neq 0$ is indeed uniform.

The function $w \mapsto \|\Sigma^{1/2}w\|_2$ Lipschitz continuous with Lipschitz constant equal to $\|\Sigma\|^{1/2}$, therefore we can apply Lévy's lemma (Lévy, 1951), again conditionally on $W \neq 0$, to get that

$$\|\Sigma^{1/2}U\|_2 \le \mathbb{E}\|\Sigma^{1/2}U\|_2 + \sqrt{\frac{2\|\Sigma\|^{1/2}t}{d}}$$

with probability at least $1-e^{-t}$ (this version of the inequality with sharp constants is due to Aubrun et al. (2024)). The same result also holds unconditionally since $U=0$ if $W=0$. It remains to observe that

$$\mathbb{E}\|\Sigma^{1/2}U\|_2 \le \mathbb{E}^{1/2}\|\Sigma^{1/2}U\|_2^2 = \sqrt{\frac{\operatorname{tr}(\Sigma)}{d}},$$

hence we conclude by the union bound that

$$\|\Sigma^{1/2}W\|_2 = \|W\|_2\|\Sigma^{1/2}U\|_2 \le e(n,t,d)\left(\sqrt{\frac{\operatorname{tr}(\Sigma)}{d}} + \sqrt{\frac{2\|\Sigma\|^{1/2}t}{d}}\right)$$

with probability at least $1-p(t)-e^{-t}$.

Remark A.1. *Spherical symmetry plays a crucial role in the proof of the theorem. Due to the affine invariance of the depth function, it is not possible to capture the dependence of the error on the effective rank by simply comparing the empirical and the true depth. For example, let $X_1,\dots,X_n$ be i.i.d. $N(0,\Sigma)$ random vectors where Σ is non-singular, and note that for any half-space $H(0,u)$, $P(X_1 \in H(0,u)) = \frac{1}{2}$. Therefore,*

$$\sup_{\|u\|_2=1}\left|\frac{1}{n}\sum_{j=1}^n I\{X_j \in H(0,u)\} - \frac{1}{2}\right| = \sup_{\|u\|_2=1}\left|\frac{1}{n}\sum_{j=1}^n I\{\Sigma^{-1/2}X_j \in H(0,u)\} - \frac{1}{2}\right|,$$

hence the distribution of the supremum does not depend on Σ.

Appendix B. Proof of Lemma 2.2.

The key step in the argument involves construction of a coupling between the empirical process

$$F_n(z,v) = \sqrt{n}(P_n^X - P)H(z,v)$$

and the Brownian bridge $G(z,v)$ indexed by the indicator functions of halfspaces. The strong approximation result by Koltchinskii (1994, Theorem 11.3) guarantees the existence of a sequence

$X_1, \ldots, X_n$ of independent standard normal random vectors and the Brownian bridge $G(v, z)$ defined on the same probability space such that

$$\sup_{z \in \mathbb{R}^d, \|v\|_2 = 1} |F_n(z, v) - G(z, v)| \le \delta_n(t) := C(d) n^{-\frac{1}{2d}} \log^{1/2}(n) \left(s + \log(n)\right) \tag{15}$$

with probability at least $1 - e^{-s}$, where $C(d)$ depends only on d. To verify that the aforementioned result indeed applies in our setting, first recall that the Vapnik-Chervonenkis dimension of half-spaces equals $d+1$ (see van der Vaart and Wellner, 2023). Next, we will represent $X_j = (x_{j,1}, \ldots, x_{j,d})$, $j = 1, \ldots, n$ as $\left(\Phi^{-1}(U_{j,1}), \ldots, \Phi^{-1}(U_{j,d})\right)$ where $U_{j,i} := \Phi(x_{j,i})$, $i = 1, \ldots, d$, $j = 1, \ldots, n$ are independent random variables with uniform distribution on $[0, 1]$. Then (15) would follow once we establish an estimate for the modulus of continuity $\omega(f_{v,b}; h)$ of the functions

$$f_{v,b}(u_1, \ldots, u_d) = I\left\{\left\langle \Phi^{-1}(\mathbf{u}), v\right\rangle - b \ge 0\right\}$$

where $\mathbf{u} := (u_1, \ldots, u_d) \in [0, 1]^d$, $\Phi^{-1}(\mathbf{u}) := \left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$, v is a fixed unit vector and $b \in \mathbb{R}$. Specifically, to deduce (15) from Theorem 11.3 in (Koltchinskii, 1994), we need to prove that for small values of h, $\omega(f_{v,b}; h) \le C\sqrt{h}$, where C does not depend on v, b and

$$\omega(f_{v,b}; h) = \sup_{\|\mathbf{r}\|_2 \le h} \left(\int_{[0,1]^d} (f_v(\mathbf{u} + \mathbf{r}) - f_v(\mathbf{u}))^2 d\mathbf{u} \right)^{1/2}.$$

Here, we assume that $f_{v,b}(u_1, \ldots, u_d) = 0$ if $u_j \notin [0, 1]$ for some j. Note that without loss of generality we can consider only the values $b \ge 0$: if $b < 0$, note that

$$I\{\langle X, v\rangle \ge b\} = 1 - I\{\langle -X, v\rangle \ge -b\}$$

where $-X$ also has standard normal distribution, and the function $1 - f_{v,b}$ has the same modulus of continuity as $f_{v,b}$. Moreover, it is enough to tackle the case $v = e_1$ where $e_1, \ldots, e_d$ is the standard Euclidean basis. Indeed, given an arbitrary v, let $v, v_2, \ldots, v_d$ be any orthonormal basis. In this basis, the coordinates $x'_{j,1}, \ldots, x'_{j,d}$ of X_j are still independent $N(0, 1)$ random variables, hence the problem reduces to $v = e_1$. Finally, we will be assuming that $r \ge 0$ (the case $r < 0$ can be handled in a similar way). Using the fact that the integrand is bounded by 1, we see that $\omega(f_{v,b}; h)$ does not exceed

$$\sup_{0 \le r \le h} \left(\int_{0 \le u_1 < 1 - r} \left(I\{\Phi^{-1}(u_1 + r) - b \ge 0\} - I\{\Phi^{-1}(u_1) - b \ge 0\} \right)^2 du_1 + \int_{\{1 - r < u_1 \le 1\}} du_1 \right)^{1/2}.$$

Since $\Phi^{-1}(u_1 + r) > \Phi^{-1}(u_1)$, the integrand in the first term above equals 1 whenever $\Phi^{-1}(u_1) < b$ and $\Phi^{-1}(u_1 + r) \ge b$, that is, for $u_1 \in [\Phi(b) - r, \Phi(b)]$, and is 0 otherwise. We conclude that

$$\omega(f_{v,b}; h) \le 2\sqrt{h}.$$

We are now ready to proceed with the proof of the main claim of the lemma. Recall the definitions (9) and (10) of $W_n(z)$ and $W(z)$ respectively, and consider two possibilities: (a) $W_n(z) - W(z) \ge 0$ and (b) $W_n(z) - W(z) < 0$. Observe that the infimum in the definitions of $W_n(z)$ and $W(z)$ is attained, almost surely, at some unit vectors $v_n(z)$ and $v(z)$. We will only consider case (a) below since the reasoning in case (b) is very similar. The following relation is evident:

$$
\begin{aligned}
(16)\quad 0 \le W_n(z) - W(z) &= \sqrt{n}\left(P_n^X H\left(\frac{z}{\sqrt{n}}, v_n(z)\right) - \frac{1}{2}\right) - \left(G(v(z),0) - \frac{1}{\sqrt{2\pi}}\langle z, v(z)\rangle\right) \\
&= \underbrace{\sqrt{n}\left(P_n^X H\left(\frac{z}{\sqrt{n}}, v_n(z)\right) - \frac{1}{2}\right) - \sqrt{n}\left(P_n^X H\left(\frac{z}{\sqrt{n}}, v(z)\right) - \frac{1}{2}\right)}_{\le 0} \\
&+ \sqrt{n}\left(P_n^X H\left(\frac{z}{\sqrt{n}}, v(z)\right) - \frac{1}{2}\right) + \frac{1}{\sqrt{2\pi}}\langle z, v(z)\rangle - \sqrt{n}(P_n^X - P)H\left(\frac{z}{\sqrt{n}}, v(z)\right) \\
&+ \sqrt{n}(P_n^X - P)H\left(\frac{z}{\sqrt{n}}, v(z)\right) - G\left(v(z), \frac{z}{\sqrt{n}}\right) + G\left(v(z), \frac{z}{\sqrt{n}}\right) - G\left(v(z), 0\right).
\end{aligned}
$$

Recalling that all 1-dimensional projections of X are standard normal, we deduce that

$$
\begin{aligned}
&\sqrt{n}\left(P_n^X H\left(\frac{z}{\sqrt{n}}, v(z)\right) - \frac{1}{2}\right) + \frac{1}{\sqrt{2\pi}}\langle z, v(z)\rangle - \sqrt{n}(P_n^X - P)H\left(\frac{z}{\sqrt{n}}, v(z)\right) \\
&= \sqrt{n}\left(PH\left(\frac{z}{\sqrt{n}}, v(z)\right) - \frac{1}{2}\right) + \frac{1}{\sqrt{2\pi}}\langle z, v(z)\rangle = \frac{1}{\sqrt{2\pi}}\langle z, v(z)\rangle - \sqrt{n}\int_0^{\left\langle \frac{z}{\sqrt{n}}, v(z)\right\rangle} \phi(t)dt \\
&\le C\frac{\|z\|_2^3}{n}
\end{aligned}
$$

where we used the fact that $\phi(t) \ge \frac{1}{\sqrt{2\pi}}(1-t^2)$ for small t. Moreover, inequality (15) implies that

$$
\sup_{z\in\mathbb{R}^d}\left|\sqrt{n}(P_n^X - P)H\left(\frac{z}{\sqrt{n}}, v(z)\right) - G\left(v(z), \frac{z}{\sqrt{n}}\right)\right| \le C(d)n^{-\frac{1}{2d}}\log^{3/2}(n)
$$

with probability at least $1 - 1/n$. Finally,

$$
\left|G\left(v(z), \frac{z}{\sqrt{n}}\right) - G\left(v(z), 0\right)\right| \le \sup_{\|v\|_2=1}\left|G\left(v, \frac{z}{\sqrt{n}}\right) - G(v,0)\right|.
$$

In view of the Gaussian concentration inequality (Tsirelson et al., 1976),

$$
\begin{aligned}
&\sup_{\|z\|_2\le R, \|v\|_2=1}\left|G\left(v, \frac{z}{\sqrt{n}}\right) - G(v,0)\right| \\
&\qquad \le \mathbb{E}\sup_{\|z\|_2\le R, \|v\|_2=1}\left|G\left(v, \frac{z}{\sqrt{n}}\right) - G(v,0)\right| + C\left(\frac{R}{n}\right)^{1/4}\sqrt{s}
\end{aligned}
$$

with probability at least $1 - e^{-s}$, where we also used (14) to estimate

$$
\mathrm{Var}\left(G\left(v, \frac{z}{\sqrt{n}}\right) - G(v,0)\right).
$$

Dudley's entropy integral bound yields, via a standard argument (as explained in the proof of Lemma 2.3), that

$$
\mathbb{E}\sup_{\|z\|_2\le R, \|v\|_2=1}\left|G\left(v, \frac{z}{\sqrt{n}}\right) - G(v,0)\right| \le K\sqrt{\frac{R}{n}}\sqrt{\log(\sqrt{n/d})}.
$$

It remains to note that for $s = \log(n)$, the last term is dominated by $C\left(\frac{R}{n}\right)^{1/4}\sqrt{s}$ since $n > C_1 R$ by assumption. In case (b), the reasoning is very similar. Collecting the estimates of each term in (16), we deduce the desired result.

B.1. **Auxiliary results.** Recall the definition (10) of the process $W(z)$. Moreover, let

$$\widehat{z} \in \arg\max_{z \in \mathbb{R}^d} W(z),$$
$$\mathcal{U} := \mathcal{U}(\widehat{z}) = \arg\min_{v:\|v\|_2=1} \left\{ G(v,0) - \frac{1}{\sqrt{2\pi}} \langle v, \widehat{z} \rangle \right\}.$$

Lemma B.1. *The following properties hold:*

(i) With probability 1, $\widehat{z}$ *is unique. Moreover, there exists event* $\mathcal{E}$ *of probability at least* $1 - e^{-t}$ *such that on this event,* $\|\widehat{z}\|_2 \le C\sqrt{d+t}$*;*

(ii) With probability 1, *for any* $w \in \mathbb{R}^d$, $w \neq 0$, *there exists* $u \in \mathcal{U}$ *such that* $u^\top w \le 0$. *Moreover, if* $d = 2$, *then the inequality can be strengthened to* $\sup_{w\neq 0} \min_{u \in \mathcal{U}} u^\top \frac{w}{\|w\|_2} < 0$.

Proof. **(i)** Lemma 3.8 in (Massé, 2002) states that with probability 1, $\arg\max_{z\in\mathbb{R}^d} W(z)$ is unique. To establish the bound on $\|\widehat{z}\|_2$, note that in view of the Gaussian concentration inequality (Tsirelson et al., 1976),

$$\inf_{\|v\|_2=1} G(v,0) \ge \mathbb{E} \inf_{\|v\|_2=1} G(v,0) - \sqrt{2t} \sup_{\|v\|_2=1} \mathrm{var}^{1/2}(G(v,0))$$

with probability at least $1-e^{-t}$. A standard argument[3] based on Dudley's entropy integral bound (Dudley, 1967) and the fact that the Vapnik-Chervonenkis dimension of the class of half-spaces passing through the origin equals d implies that

$$\mathbb{E} \inf_{\|v\|_2=1} G(v,0) \ge -C_1\sqrt{d}.$$

Since $\mathrm{var}(G(v,0)) = 1/4$ for all v, we deduce that

$$\inf_{\|v\|_2=1} G(v,0) \ge -C_2\sqrt{d+t}$$

on event $\mathcal{E}$ of probability at least $1 - e^{-t}$. We conclude that on $\mathcal{E}$,

$$W(0) = \inf_{\|v\|_2=1} G(v,0) \ge -C_2\sqrt{d+t}.$$

Moreover, for $\|z\|_2 > C\sqrt{d+t}$, we have

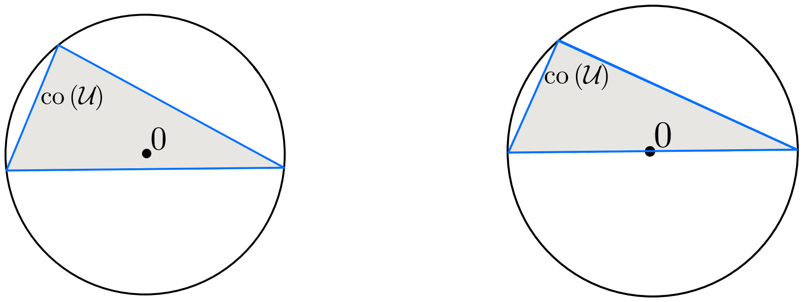
$$W(z) \le -\frac{1}{\sqrt{2\pi}}\|z\|_2 + \sup_{\|v\|_2=1} G(v,0) < C_2\sqrt{d+t},$$

whenever C is sufficiently large. Therefore, $\|\widehat{z}\|_2 \le C\sqrt{d+t}$ on $\mathcal{E}$.

(ii) The first claim is established in the course of the proof of Lemma 3.8 in (Massé, 2002), also see (Nolan, 1999). Assume now that $d = 2$. Massé (2002); Nolan (1999) show that the set $\mathcal{U}$ must contain at least 3 elements. Trajectories of the process $W(z)$ are concave with probability 1 as the pointwise infimum of linear functions, moreover, the subdifferential $\partial W(\widehat{z})$ is the convex hull $\mathrm{co}(\mathcal{U})$ of the set $\mathcal{U}$ (Boyd and Vandenberghe, 2004). The necessary condition for $\widehat{z}$ to be the maximizer of $W(z)$ is $0 \in \partial W(\widehat{z})$. If 0 is on the boundary of $\mathrm{co}(\mathcal{U})$, then there exists a chord passing through 0 contained in $\mathrm{co}(\mathcal{U})$, implying that there exists a unit vector $\hat{v}$ such that $\{\hat{v}, -\hat{v}\} \subset \mathcal{U}$ (see figure 4b). But in this case

$$M(\widehat{z}, u) := G(u,0) - \frac{1}{\sqrt{2\pi}}\langle \widehat{z}, u\rangle = 0 \tag{17}$$

[3]For example, it suffices to combine Theorem 2.37 with Theorem 4.47 or 4.52 in (Dudley, 2014).



(A) Depiction of the case when 0 is in the interior of co$(\mathcal{U})$.

(B) Depiction of the case when 0 is on the boundary of co$(\mathcal{U})$.

FIGURE 4. Subdifferential of $W(\widehat{z})$.

for all unit vectors u, in which case $\mathcal{U} = S^2$ and the claim readily holds. To verify (17), observe that, since $G(-u,0) = -G(u,0)$ with probability 1, $M(z,-u) = -M(z,u)$ for all z, u. Therefore,

$$W(\widehat{z}) = M(\widehat{z},\hat{v}) = \inf_{\|u\|_2=1} M(\widehat{z},u) \le \sup_{\|u\|_2=1} M(\widehat{z},u) = M(\widehat{z},-\hat{v}) = W(\widehat{z}),$$

implying the claim.
Finally, when 0 is in the interior of co$(\mathcal{U})$ (figure 4a), then

$$\sup_{w\neq 0} \min_{u\in\mathcal{U}} u^\top \frac{w}{\|w\|_2} = -\mathrm{dist}\left(0, \mathrm{bd}(\mathrm{co}(\mathcal{U}))\right) < 0$$

where $\mathrm{dist}(x, A)$ is the distance from a point x to a set A and $\mathrm{bd}(\mathrm{co}(\mathcal{U}))$ is the boundary of co$(\mathcal{U})$. □

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTHERN CALIFORNIA
Email address: minsker@usc.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SOUTHERN CALIFORNIA
Email address: yinanshe@usc.edu