THE PROJECTION MEDIAN AS A WEIGHTED AVERAGE*

Stephane Durocher,[†] *Alexandre Leblanc,*[†] *and Matthew Skala*[‡]

Abstract. The projection median of a set P of n points in $\mathbb{R}^d$ is a robust geometric generalization of the notion of univariate median to higher dimensions. In its original definition, the projection median is expressed as a normalized integral of the medians of the projections of P onto all lines through the origin. We introduce a new definition in which the projection median is expressed as a weighted mean of P, and show the equivalence of the two definitions. In addition to providing a definition whose form is more consistent with those of traditional statistical estimators of location, this new definition for the projection median allows many of its geometric properties to be established more easily, as well as enabling new randomized algorithms that compute approximations of the projection median with increased accuracy and efficiency, reducing computation time from $O(n^{d+\epsilon})$ to $O(mnd)$, where m denotes the number of random projections sampled. Selecting $m \in \Theta(\epsilon^{-2}d^2 \log n)$ or $m \in \Theta(\min(d + \epsilon^{-2} \log n, \epsilon^{-2}n))$ suffices for our algorithms to return a point within relative distance ϵ of the true projection median with high probability, resulting in running times $O(d^3 n \log n)$ and $O(\min(d^2 n, dn^2))$ respectively, for any fixed ϵ.

1 Introduction

The median is a fundamental statistic in data analysis. Given a multiset P of n elements drawn from a totally ordered universe U, a (univariate) *median* of P is any value x in U that lies in the interval $[p_{\lfloor n/2 \rfloor}, p_{\lceil n/2 \rceil}]$, where p_i denotes the element of rank i in P, for each i in $\{1, \ldots, n\}$. When $U = \mathbb{R}$, it is straightforward to confirm that x is a balance point that minimizes the sum of the distances (equivalently, the average distance) from x to the elements of P. The one-dimensional notion of median can be generalized to higher dimensions in multiple ways. One natural approach is to consider the *Weber point* (also known as Fermat point, Torricelli point, geometric median, or Euclidean median) of a multiset of points $P \subseteq \mathbb{R}^d$: a point in $\mathbb{R}^d$ (not necessarily in P) that minimizes the sum of the Euclidean distances from itself to points in P. The Weber point is highly unstable [17]. That is, an arbitrarily small perturbation of the points in P can result in an arbitrarily large relative change in the Weber point of P. This instability can lead to inconsistent or incorrect conclusions being drawn during data analysis. Furthermore, no algorithm is known for exact computation of the Weber point [4, 19]. The only known solutions are algorithms that compute an approximation to the Weber point. These include Weizfeld's

*This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

† *University of Manitoba*, durocher@cs.umanitoba.ca, alex_leblanc@umanitoba.ca

‡ *North Coast Synthesis Ltd.*, mskala@ansuz.sooke.bc.ca

algorithm [29] and more recent approximation algorithms [6, 9, 20]. Like the exact Weber point, however, these previous approximations are highly unstable. See Kupitz and Martini [21] or Wesolowsky [30] for more discussion of the Weber point and its properties.

Alternatively, Durocher and Kirkpatrick [17] introduced the projection median, denoted $\mathcal{M}_{\mathrm{P}}(P)$, of a set of points P in $\mathbb{R}^2$ as a stable geometric generalization of the notion of univariate median to higher dimensions for which the sum of the distances from $\mathcal{M}_{\mathrm{P}}(P)$ to points in P closely approximates the minimum sum achieved by the Weber point (see Section 5) and whose position can be computed exactly (see Section 3.1). This definition generalizes to $\mathbb{R}^d$ [14]. The examination of the projection median was continued by Basu et al. [5], who established and extended several of its properties to higher dimensions. In its original definition (see (1) in Section 2) the projection median of a set P of n points in $\mathbb{R}^d$ is expressed as a normalized integral of the medians of the projections of P onto all lines through the origin. This definition is similar in structure to that of the Steiner centre of P, defined as a normalized integral of the extrema of the projections of P onto all lines through the origin. Shephard [27] showed an equivalent expression for the Steiner centre as a weighted mean of P. These dual expressions for the Steiner centre allow various geometric, statistical, and algorithmic properties to be established for it [16]. This paper seeks to identify a similar expression for the projection median as a weighted mean, with the objective to apply the new expression to develop more efficient algorithms, to establish new properties, and to expand the potential applications of the projection median by expressing it in a form more traditionally associated with statistical estimators of location. We summarize our contributions below.

Expressing the Projection Median as a Weighted Mean. In Section 2, we introduce an expression for the projection median as a weighted average (2) and show its equivalence to the original definition of the projection median (1). The equivalence is non-trivial and the techniques employed differ from those used by Shephard [27] for the Steiner centre. We generalize this result to express the equivalence of the weighted mean and integral formulations of a range of summary statistics defined as functions of order statistics, including the projection median at one end, and the Steiner centre at the other.

Efficient Approximation Algorithms. The two expressions for the projection median, (1) and (2), suggest randomized algorithms to compute $\mathcal{M}_{\mathrm{P}}(P)$ using a discrete set of projections selected at random. In Section 3, we introduce two randomized algorithms, each defined in terms of one of the expressions for the projection median, and analyze these to establish probabilistic bounds on the quality of the approximations returned. Our randomized algorithms are easy to implement and run in $O(mnd)$ time, where m denotes the number of random projections selected, improving on the time of $O(n^{d+\epsilon})$ required by the current most efficient algorithm for computing the projection median exactly [5]. All points in P are included in the computation of the approximation; the approximation is achieved by randomly selecting a discrete set of unit vectors, as opposed to selecting a random subset of P. In Lemmas 4 and 7 we bound the probabilities that the approximate projection medians returned by our algorithms lie within relative[1] distance ϵ of the true

[1]The relative error is defined by $\epsilon = \varepsilon/(\max_{p\in P} \|p\|_2)$, where ε denotes the difference in the positions of the approximate and true projection medians. See Section 3 for details.

projection median as functions of m, n, and d. In particular, selecting $m \in \Omega(\epsilon^{-2} d^2 \log n)$ or $m \in \Omega(\min(d + \epsilon^{-2} \log n, \epsilon^{-2} n))$ ensures that each algorithm returns a point within relative distance ϵ of the true projection median with high probability, i.e., with probability at least $1 - n^{-a}$ for some constant $a \geq 1$. The resulting running times are $O(d^3 n \log n)$ and $O(\min(d^2 n, dn^2))$ respectively, for any fixed ϵ. Furthermore, we show that the probability of closeness approaches 1 exponentially quickly as a function of m. According to these bounds, Algorithm 2, defined in terms of the weighted mean expression for the projection median (2), is expected to require smaller m (directly proportional to running time) than Algorithm 1 to achieve the same precision for any given n and d, i.e., $O(d^2 \log n)$ for Algorithm 1 versus $O(\min(d, n))$ for Algorithm 2. Furthermore, the accuracy of Algorithm 2 is independent of the location of the origin relative to the true projection median; as we show in Section 4, the accuracy of Algorithm 1 depends on the true projection median's distance to the origin. Thus, Algorithm 2 is robust with respect to similarity transformations, whereas Algorithm 1 is not, further motivating the new expression for the projection median as a weighted mean. Definitions and theoretical analysis for both algorithms are given in Section 3. Experimental analysis and discussion are found in Section 4.

Geometric Properties. In Section 5, we use the new expression of the projection median as a weighted mean to provide arguments for establishing several of its geometric properties, including equivariance under similarity transformations and consistency across dimensions. These properties are compared against those of other multivariate estimators of median in Section 5.2.

2 Expressing the Projection Median as a Weighted Mean

Given a multiset of points P in $\mathbb{R}^d$ and a unit vector $u \in S^{d-1}$, where $S^{d-1} = \{u \in \mathbb{R}^d \mid \|u\| = 1\}$ is the d-dimensional unit hypersphere centred at the origin, let l_u denote the directed line in $\mathbb{R}^d$ through the origin in the direction u and let $P_u = \{u\langle p, u\rangle \mid p \in P\}$ denote the multiset determined by the projection of P onto l_u. Durocher and Kirkpatrick [17] introduced the projection median of a multiset of points P in $\mathbb{R}^d$ as a generalization of the one-dimensional median to higher dimensions, defined as

$$\mathcal{M}_{\mathrm{P}}(P) = d \frac{\int_{S^{d-1}} \mathrm{med}(P_u)\ \mathrm{d}u}{\int_{S^{d-1}} 1\ \mathrm{d}u}, \tag{1}$$

where $\mathrm{med}(P_u)$ is a one-dimensional median of P_u. We use non-italic d for the differential to distinguish it from the number of dimensions d.

If $|P|$ is even, then the median of P_u may not be uniquely defined. For each u and corresponding $\mathrm{med}(P_u)$, the computation of $\mathcal{M}_{\mathrm{P}}(P)$ also includes $-u$ and $\mathrm{med}(P_{-u})$. Consequently, if the definition of $\mathrm{med}(P_u)$ is consistent relative to u, then $\mathrm{med}(P_u)$ and $\mathrm{med}(P_{-u})$ are averaged, giving the midpoint of the interval of medians in $P_u = P_{-u}$. Therefore, $\mathcal{M}_{\mathrm{P}}(P)$ is uniquely defined regardless of the median of P_u chosen, so long as the choice is consistent for P_{-u}. Without loss of generality, suppose $\mathrm{med}(P_u)$ denotes the leftmost (with respect to u) median of P_u. Finally, let $p_*(u)$ denote the point in P whose projection determines the median of P_u, i.e., the point $p \in P$ such that $u\langle p, u\rangle = \mathrm{med}(P_u)$. See Figure 1 for examples in $\mathbb{R}^2$.

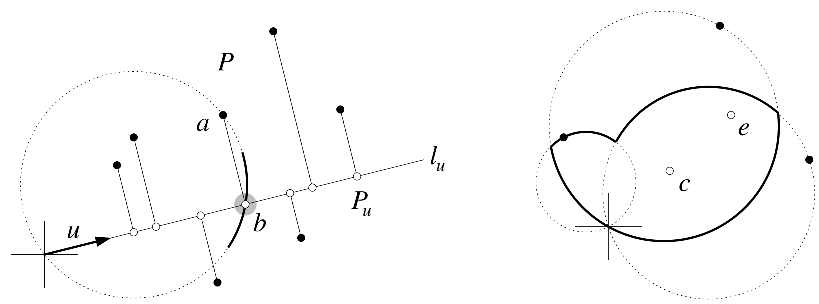


Figure 1: **Left:** A set P (seven black points) in $\mathbb{R}^2$, a unit vector u, the line l_u through the origin determined by u, the multiset P_u (white points) determined by projecting P onto l_u, and the point b (shaded) which is the projection of a onto l_u and the median of P_u, denoted $\text{med}(P_u)$. Note that $p_*(u) = a$. As u rotates about the origin, the change in position of b follows an arc (bold) of the circle tangent to the origin, a, and b until $a \neq p_*(u)$. **Right:** Given a second set P (three black points) in $\mathbb{R}^2$, as u rotates about the origin, the position of $\text{med}(P_u)$ follows a continuous trajectory along arcs (bold) of circles. Integrating $\text{med}(P_u)$ gives a point c that is a weighted mean of the arcs. Scaling by a factor of d (here $d = 2$) gives a point e within the convex hull of P that defines the projection median of P, denoted $\mathcal{M}_{\text{P}}(P)$.

As we show in Theorems 1 and 2, the projection median of a finite multiset of points P in $\mathbb{R}^d$ as defined in (1) can also be expressed as

$$\mathcal{M}_{\text{A}}(P) = \sum_{p \in P} w_p p, \tag{2}$$

where for each $p \in P$, $w_p \in [0, 1]$ denotes the proportion of the unit vectors $u \in S^{d-1}$ such that the projection of p onto l_u is a median of P_u, i.e., such that $p = p_*(u)$. Formally, for each $p \in P$,

$$w_p = \frac{\int_{S_p^{d-1}} 1 \, \mathrm{d}u}{\int_{S^{d-1}} 1 \, \mathrm{d}u}, \quad \text{where} \quad S_p^{d-1} = \{u \mid u \in S^{d-1} \text{ and } p = p_*(u)\}. \tag{3}$$

The sets S_p^{d-1} for $p \in P$ partition S^{d-1}. Consequently,

$$\begin{aligned}
\sum_{p \in P} w_p &= \frac{\sum_{p \in P} \int_{S_p^{d-1}} 1 \, \mathrm{d}u}{\int_{S^{d-1}} 1 \, \mathrm{d}u} \\
&= \frac{\int_{\cup_{p \in P} S_p^{d-1}} 1 \, \mathrm{d}u}{\int_{S^{d-1}} 1 \, \mathrm{d}u} \\
&= \frac{\int_{S^{d-1}} 1 \, \mathrm{d}u}{\int_{S^{d-1}} 1 \, \mathrm{d}u} \\
&= 1\,.
\end{aligned}$$

2.1 Equivalence in Two Dimensions

We now show the equivalence of the two definitions of the projection median, (1) and (2). The special case of two dimensions is important in itself, and provides intuition for the general case, so we give a specific proof for two dimensions first.

Theorem 1. *For any finite multiset $P \subseteq \mathbb{R}^2$, $\mathcal{M}_{\mathrm{P}}(P) = \mathcal{M}_{\mathrm{A}}(P)$.*

Proof. Choose any finite multiset P containing n elements, each of which is a point in $\mathbb{R}^2$. Let $q_0 = (x_0, y_0), \ldots, q_{k-1} = (x_{k-1}, y_{k-1})$ denote the sequence of k points in P whose projections are medians of P_θ as θ varies over $[0, 2\pi)$, where for each i, $[\theta_i, \theta_{i+1}) \subseteq [0, 2\pi)$ denotes the interval associated with q_i. That is, for each i and each $\theta \in [\theta_i, \theta_{i+1})$, $q_i = p_*(\theta)$. A point in P may realize a median in P_θ multiple times as θ varies, meaning that two points, q_i and q_j, in the sequence could be equal, where $i \neq j$. Since the projection of each point q_j in P_θ changes continuously with θ, when the projection of q_i ceases to be a median of P_θ (at $\theta = \theta_{i+1}$), the projections of q_i and q_{i+1} coincide on $l_{\theta_{i+1}}$. Therefore,

$$\begin{aligned} \langle u_{\theta_{i+1}}, q_i \rangle &= \langle u_{\theta_{i+1}}, q_{i+1} \rangle \\ \Rightarrow x_i \cos\theta_{i+1} + y_i \sin\theta_{i+1} &= x_{i+1}\cos\theta_{i+1} + y_{i+1}\sin\theta_{i+1}, \end{aligned} \tag{4}$$

where the unit vector $u_\theta = (\cos\theta, \sin\theta)$. By definition,

$$\mathcal{M}_{\mathrm{P}}(P) = \frac{2}{2\pi}\int_0^{2\pi} \mathrm{med}(P_\theta)\ \mathrm{d}\theta = \frac{1}{\pi}\sum_{i=0}^{k-1}\int_{\theta_i}^{\theta_{i+1}} u_\theta \langle q_i, u_\theta\rangle\ \mathrm{d}\theta. \tag{5}$$

Now consider only the x-coordinate in (5):

$$\begin{aligned} \mathcal{M}_{\mathrm{P}}(P)_x &= \frac{1}{\pi}\sum_{i=0}^{k-1}\int_{\theta_i}^{\theta_{i+1}} \cos\theta(y_i\sin\theta + x_i\cos\theta)\ \mathrm{d}\theta \\ &= \frac{1}{2\pi}\sum_{i=0}^{k-1}\left[-y_i\cos^2\theta + x_i(\sin\theta\cos\theta + \theta)\right]\Big|_{\theta_i}^{\theta_{i+1}} \\ &= \frac{1}{2\pi}\sum_{i=0}^{k-1} x_i(\theta_{i+1}-\theta_i) + \sin\theta_{i+1}(y_i\sin\theta_{i+1} + x_i\cos\theta_{i+1}) - \sin\theta_i(y_i\sin\theta_i + x_i\cos\theta_i) \\ &= \frac{1}{2\pi}\sum_{i=0}^{k-1} x_i(\theta_{i+1}-\theta_i) + \sin\theta_{i+1}(y_{i+1}\sin\theta_{i+1} + x_{i+1}\cos\theta_{i+1}) \\ &\quad - \sin\theta_i(y_i\sin\theta_i + x_i\cos\theta_i) \qquad \text{(by (4))} \\ &= \frac{1}{2\pi}\left[\sum_{i=0}^{k-1} x_i(\theta_{i+1}-\theta_i) + \sum_{i=0}^{k-1}\sin\theta_{i+1}(y_{i+1}\sin\theta_{i+1} + x_{i+1}\cos\theta_{i+1}) \right. \\ &\quad \left. - \sum_{i=0}^{k-1}\sin\theta_i(y_i\sin\theta_i + x_i\cos\theta_i)\right], \end{aligned}$$

and since all indices are modulo k (e.g., $\theta_{j+k} = \theta_j$),

$$\begin{aligned}
\mathcal{M}_{\mathrm{P}}(P)_x &= \frac{1}{2\pi}\left[\sum_{i=0}^{k-1} x_i(\theta_{i+1}-\theta_i) + \sum_{i=0}^{k-1} \sin\theta_i(y_i\sin\theta_i + x_i\cos\theta_i)\right. \\
&\qquad \left. - \sum_{i=0}^{k-1} \sin\theta_i(y_i\sin\theta_i + x_i\cos\theta_i)\right] \\
&= \frac{1}{2\pi}\sum_{i=0}^{k-1} x_i(\theta_{i+1}-\theta_i) \\
&= \sum_{i=0}^{k-1} w_{q_i} x_i\,, \qquad (6)
\end{aligned}$$

where $w_{q_i} = (\theta_{i+1}-\theta_i)/(2\pi)$. By (6) and an analogous argument on the y-coordinate of (5),

$$\mathcal{M}_{\mathrm{P}}(P) = \sum_{i=0}^{k-1} w_{q_i} q_i = \sum_{p\in P} w_p p = \mathcal{M}_{\mathrm{A}}(P)\,. \qquad \square$$

2.2 Equivalence in Higher Dimensions

Now we show that the equivalence holds in arbitrary dimension.

Theorem 2. *For any finite multiset $P \subseteq \mathbb{R}^d$, $\mathcal{M}_{\mathrm{P}}(P) = \mathcal{M}_{\mathrm{A}}(P)$.*

Proof. The result follows trivially when $d = 1$. Therefore, choose any $d \geq 2$. A d-dimensional unit vector $u \in S^{d-1}$ can be described by generalized spherical coordinates $\phi_1, \phi_2, \ldots, \phi_{d-1}$, where each ϕ_i ranges from 0 to π, except ϕ_{d-1} which ranges from 0 to 2π. The Cartesian coordinates of u are $u_1 = \cos\phi_1$, $u_2 = \sin\phi_1\cos\phi_2$, $\ldots$, $u_{d-1} = \sin\phi_1\cdots\sin\phi_{d-2}\cos\phi_{d-1}$, and $u_d = \sin\phi_1\cdots\sin\phi_{d-2}\sin\phi_{d-1}$. The area element $\mathrm{d}u$, which can be thought of as an infinitesimal patch of $(d-1)$ dimensional space tangent to the surface of the sphere, is given by $\mathrm{d}u = \sin^{d-2}\phi_1\sin^{d-3}\phi_2\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_1\mathrm{d}\phi_2\cdots\mathrm{d}\phi_{d-1}$. Let A be the generalized surface area of S^{d-1}. It can be calculated by integrating $\mathrm{d}u$:

$$\begin{aligned}
A &= \int_{S^{d-1}} 1\ \mathrm{d}u \\
&= \int_{\phi_{d-1}}\int_{\phi_{d-2}}\cdots\int_{\phi_1} \sin^{d-2}\phi_1\sin^{d-3}\phi_2\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_1\mathrm{d}\phi_2\cdots\mathrm{d}\phi_{d-1} \\
&= \frac{2\pi^{d/2}}{\Gamma(d/2)}\,.
\end{aligned}$$

While we don't require the exact value of A for our purposes, we note that after increasing to a maximum for integer d when $d = 7$, the generalized surface area approaches zero as d goes to infinity.

From the definitions of $\mathcal{M}_{\mathrm{A}}(P)$ and w_p we have

$$\mathcal{M}_{\mathrm{A}}(P) = \frac{1}{A}\int_{S^{d-1}} 1\cdot p_*(u)\,\mathrm{d}u\,;$$

and because $\mathrm{med}(P_u) = \langle p_*(u), u\rangle u$,

$$\mathcal{M}_{\mathrm{P}}(P) = \frac{d}{A}\int_{S^{d-1}} \langle p_*(u), u\rangle u \,\mathrm{d}u\,. \tag{7}$$

Splitting (7) into a sum over all the sets S_p^{d-1}, we get

$$\mathcal{M}_{\mathrm{P}}(P) = \frac{d}{A}\sum_{p\in P}\int_{S_p^{d-1}} \langle p, u\rangle u \,\mathrm{d}u\,, \tag{8}$$

which is an equation in d-dimensional vectors. It can equivalently be treated as a set of d equations in real numbers, one for each dimension.

Choose any dimension, say without loss of generality the first, and let $\mathcal{M}_{\mathrm{P}}(P)_1$ denote the corresponding coordinate of $\mathcal{M}_{\mathrm{P}}(P)$. Since the definitions of $\mathcal{M}_{\mathrm{P}}(P)$ and $\mathcal{M}_{\mathrm{A}}(P)$ are coordinate-free, we are free to choose the generalized spherical coordinates such that the unit vector corresponding to all $\phi_i = 0$ is the unit vector in our chosen dimension, and $u_1 = \cos\phi_1$ regardless of which dimension we chose to represent by u_1. We have

$$\mathcal{M}_{\mathrm{P}}(P)_1 = \frac{d}{A}\sum_{p\in P}\int_{\phi_{1\cdots d-1}} \langle p, u\rangle \cos\phi_1 \sin^{d-2}\phi_1 \sin^{d-3}\phi_2 \cdots \sin\phi_{d-1}\,\mathrm{d}\phi_1\mathrm{d}\phi_2\cdots\mathrm{d}\phi_{d-1}\,.$$

We now apply the integration by parts formula $\int x\,\mathrm{d}y = xy - \int y\,\mathrm{d}x$ to the integration over ϕ_1. Let

$$\begin{aligned}
\mathrm{d}y &= \cos\phi_1 \sin^{d-2}\phi_1\,\mathrm{d}\phi_1\ ,\\
y &= \frac{1}{d-1}\sin^{d-1}\phi_1\ ,\\
x &= \langle p, u\rangle \sin^{d-3}\phi_2\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_2\cdots\mathrm{d}\phi_{d-1}\\
&= [p_1\cos\phi_1 + (\sin\phi_1)\cdot(p_2\cos\phi_2+\cdots)]\cdot\sin^{d-3}\phi_2\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_2\cdots\mathrm{d}\phi_{d-1}\ ,\\
\mathrm{d}x &= [-p_1\sin\phi_1 + (\cos\phi_1)\cdot(p_2\cos\phi_2+\cdots)]\cdot\sin^{d-3}\phi_2\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_1\cdots\mathrm{d}\phi_{d-1}\ .
\end{aligned} \tag{9}$$

These values are chosen so that

$$\mathcal{M}_{\mathrm{P}}(P)_1 = \frac{d}{A}\sum_{p\in P}\int_{\phi_{1\cdots d-1}} x\,\mathrm{d}y\,.$$

But from (9) and the identity $\sin^2 = 1-\cos^2$, we have

$$\begin{aligned}
x\,\mathrm{d}y &= \langle p, u\rangle\cos\phi_1\sin^{d-2}\phi_1\sin^{d-3}\phi_2\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_1\mathrm{d}\phi_2\cdots\mathrm{d}\phi_{d-1}\\
y\,\mathrm{d}x &= \frac{1}{d-1}\Big[-p_1\sin^2\phi_1 + (\sin\phi_1\cos\phi_1)\cdot(p_2\cos\phi_2+\cdots)\Big]\\
&\qquad\cdot\sin^{d-2}\phi_1\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_1\cdots\mathrm{d}\phi_{d-1}\\
&= \frac{1}{d-1}\left[-p_1 + p_1\cos_2\phi_1 + (\sin\phi_1\cos\phi_1)\cdot(p_2\cos\phi_2+\cdots)\right]\\
&\qquad\cdot\sin^{d-2}\phi_1\cdots\sin\phi_{d-1}\,\mathrm{d}\phi_1\cdots\mathrm{d}\phi_{d-1}\\
&= \frac{1}{d-1}\left[-p_1\,\mathrm{d}u + x\,\mathrm{d}y\right]\ .
\end{aligned}$$

Applying the integration by parts formula and re-arranging,

$$\begin{aligned}\int_{\phi_{1\cdots d-1}} x\,\mathrm{d}y &= \int_{\phi_{2\cdots d-1}} \left[xy|_{\phi_1} - \int_{\phi_1} y\,\mathrm{d}x \right] \\ &= \int_{\phi_{2\cdots d-1}} \left[xy|_{\phi_1} + \frac{p_1}{d-1}\int_{\phi_1} \mathrm{d}u - \frac{1}{d-1}\int_{\phi_1} x\,\mathrm{d}y \right] \\ &= \frac{p_1}{d}\int_{S_p^{d-1}} \mathrm{d}u + \frac{d-1}{d}\int_{\phi_{2\cdots d-1}} xy|_{\phi_1}\ .\end{aligned}$$

The notation $xy|_{\phi_1}$ denotes that xy is a function of ϕ_1 and that, if we are integrating over the range of ϕ_1 from a to b, then $xy|_{\phi_1} = xy(b) - xy(a)$. Substituting this into (8), using the definition of $p_*(u)$ to combine the sum over p into an integral over all of S^{d-1} and using the definitions of x and y, we have

$$\begin{aligned}\mathcal{M}_{\mathrm{P}}(P)_1 &= \frac{1}{A}\int_{S^{d-1}} p_*(u)_1\,\mathrm{d}u + \frac{d-1}{A}\sum_{p\in P}\int_{\phi_{2\cdots d-1}} xy|_{\phi_1} \\ &= \frac{1}{A}\int_{S^{d-1}} p_*(u)_1\,\mathrm{d}u + \frac{1}{A}\sum_{p\in P}\int_{\phi_{2\cdots d-1}} \langle p,u\rangle \sin\phi_1 \frac{\mathrm{d}u}{\mathrm{d}\phi_1}\bigg|_{\phi_1}\ .\end{aligned}$$

For any fixed $\phi_2 \cdots \phi_{d-1}$, the terms in the sum over p correspond to a set of intervals of values of ϕ_1, disjoint except at their endpoints, that cover the range from $\phi_1 = 0$ to $\phi_1 = \pi$. As in the proof of Theorem 1, let $q_0, \ldots, q_k$ be the sequence of points from P that are medians of P projected onto u as ϕ_1 goes from 0 to π. Let $\phi_{1,i}$ be the value of ϕ_1 at which q_i becomes the median. Moving the summation inside the integral and writing u as a function of ϕ_1, the integrand becomes

$$\begin{aligned}&[-\langle q_0, u(0)\rangle \sin 0 + \langle q_0, u(\phi_{1,1})\rangle \sin\phi_{1,1} \\ &\qquad - \langle q_1, u(\phi_{1,1})\rangle \sin\phi_{1,1} + \langle q_1, u(\phi_{1,2})\rangle \sin\phi_{1,2} \cdots + \langle q_k, u(\pi)\rangle \sin\pi]\frac{\mathrm{d}u}{\mathrm{d}\phi_1}\,.\end{aligned}$$

The first and last terms are zero because $\sin 0 = \sin\pi = 0$. The remaining terms telescope because $\langle q_i, u(\phi_{1,i+1})\rangle = \langle q_{i+1}, u(\phi_{1,i+1})\rangle$. The sum and the integral vanish. Recombining the per-dimension equations and applying (2.2), we are left with

$$\mathcal{M}_{\mathrm{P}}(P) = \frac{1}{A}\int_{S^{d-1}} p_*(u)\,\mathrm{d}u = \mathcal{M}_{\mathrm{A}}(P)\,. \qquad \square$$

This proof may seem more complicated than necessary, especially when compared to the proof due to Shephard [27] of a similar result for the Steiner centre using coordinate-free techniques. Unfortunately, his approach requires K (equivalent to our P) to be a smooth convex set, so that the "point of contact" function $\xi(\mathbf{u}, K)$ can be smooth, continuous, and topologically well-behaved. Even in his setting, since K is not actually smooth, it is necessary to invoke a sequence of smooth sets uniformly converging on the original K. Smoothness and continuity are more difficult in our problem because we consider individual points instead of only their convex hull; the value of $p_*(u)$ jumps not only non-smoothly, but also discontinuously, from one point in P to the next. Smoothness might be addressed as Shephard does with a uniformly converging sequence, but there is also no obvious way to achieve convexity at all, so the convexity requirement would have to be eliminated; and then topological difficulties arise. A proof of this type may still be possible, but would not be simpler. Our coordinate-based approach has the advantage of requiring only elementary calculus.

Basu et al. [5] generalized a 2-dimensional result of Durocher and Kirkpatrick [17] to show that in arbitrary dimension, the rectilinear median integrated over all rotations of the coordinate system is equivalent to the projection median. Our Theorems 1 and 2 could also be written in terms of that equivalence. Any one coordinate of the rectilinear median, integrated over all rotations in $SO(d)$ with the Haar measure, is essentially equivalent to the one-dimensional projected median integrated uniformly over all unit vectors in S^{d-1}. Using the equivalence to rectilinear median avoids explicit calculation of the area element in polar coordinates, and might be advantageous if we were also proving other results about the rectilinear median; but integrating over the much more complicated object $SO(d)$ instead of S^{d-1} requires significant additional machinery and notation. We prefer to use the definition of projection median directly. The direct approach also simplifies the generalization to other order statistics in Section 2.3.

2.3 Other Order Statistics

Nothing in the proof of Theorem 2 actually requires that $\text{med}(P_u)$ is the median of P_u in particular. Any other function of u that defines a sequence of points $q_0, \ldots, q_k \in P$ with the property $\langle q_i, u(\phi_{1,i+1})\rangle = \langle q_{i+1}, u(\phi_{1,i+1})\rangle$ would suffice; for instance, any other order statistic of P_u. Let $P_u^{(i)}$ be the i-th order statistic of P projected onto u. That is, $\langle P_u^{(1)}, u\rangle \leq \langle P_u^{(2)}, u\rangle \leq \cdots \leq \langle P_u^{(n)}, u\rangle$; and let $w_p^{(i)}$ be the proportion of the unit vectors $u \in S^{d-1}$ such that the projection of p onto l_u is $P_u^{(i)}$. For any integer $1 \leq i \leq n$ we can define a projected i-th order statistic in both integral and weighted mean versions, analogous to $\mathcal{M}_\text{P}(P)$ and $\mathcal{M}_\text{A}(P)$:

$$\mathcal{M}_\text{P}^{(i)}(P) = d\frac{\int_{S^{d-1}} P_u^{(i)}\ \mathrm{d}u}{\int_{S^{d-1}} 1\ \mathrm{d}u}$$
$$\mathcal{M}_\text{A}^{(i)}(P) = \sum_{p\in P} w_p^{(i)} p\,.$$

Appropriate substitutions in the proof of Theorem 2 give the following corollary.

Corollary 3. *For any finite multiset $P \subseteq \mathbb{R}^d$ and any integer $1 \leq i \leq |P|$, $\mathcal{M}_\text{P}^{(i)}(P) = \mathcal{M}_\text{A}^{(i)}(P)$.*

Shephard's result for the Steiner centre corresponds to the special cases $i = 1$ and $i = |P|$ [27]. Corollary 3 also holds for any linear combination of order statistics. For instance, the *trimmed mean* found by taking the mean of all elements within an interval of order statistics, such as from the first to the ninth decile, is sometimes used as a statistical measure of central tendency. A trimmed mean has some of the properties of the mean, but may be less sensitive to outliers. We can generalize the trimmed mean to higher dimensions by defining a projection version of it. Then, because any trimmed mean is a linear combination of order statistics, Corollary 3 applies and the projection trimmed mean is a weighted mean of the points in P. The weight for each point corresponds to the fraction of unit vectors for which that point, when projected, is included in the one-dimensional trimmed mean.

3 Randomized Algorithms for Computing the Projection Median

In this section we define and analyze two randomized algorithms that approximate the two expressions for the projection median, (1) and (2). Both algorithms involve selecting a sequence of unit vectors $\{u_1, \dots, u_m\} \subseteq S^{d-1}$ at random. Algorithm 1 computes a normalized average of the medians of $P_{u_1}, \dots, P_{u_m}$ to approximate (1). Algorithm 2 estimates the weight of each point $p \in P$ by counting the number of random unit vectors u_i for which $p = p_*(u_i)$; the corresponding weighted mean approximates (2).

3.1 Computing the Projection Median Exactly in $O(n^{d+\epsilon})$ Time

Although expressing the projection median as a weighted mean rather than as an integral of projections can help simplify its computation, brute-force algorithms for computing the projection median exactly using the two definitions have identical asymptotic running times in the worst case. The order of the points of P projected onto a line l_u is a permutation $\pi(u)$ and this permutation determines the point $p_*(u)$ associated with $\mathrm{med}(P_u)$. The brute-force algorithms require partitioning the unit sphere S^{d-1} into regions $R_1, \dots, R_k$, each of which is the preimage under π of some permutation. If we consider S^{d-1} as a $(d-1)$-dimensional space cut by $\binom{n}{2}$ $(d-2)$-dimensional hyperplanes, each normal to the line determined by one of the $\binom{n}{2}$ pairs of points in P, we can treat one of the cuts as the equator of the sphere, project from the centre onto two hyperplanes tangent to the poles, and find that the number of regions on S^{d-1} is exactly twice the number of regions in a $(d-1)$-dimensional Euclidean space cut by $\binom{n}{2}-1$ hyperplanes. It then follows from a result of Price [23] that the number of regions for P in general position is $\Theta(n^{2d-2})$ (he also gives an exact recurrence relation for it), and that remains an upper bound with degenerate P. The number of regions is a lower bound on the worst-case time of the brute-force algorithms.

Since it suffices to compute regions in which the median's identity remains unchanged, only a subset of the $\Theta(n^{2d-2})$ regions is actually required. An algorithm only needs to identify boundaries of regions at which the identity of the median changes. The problem can be recast as a median-level computation in an arrangement of n hyperplanes. See Basu et al. [5] for a detailed description of the dual transformation. In $\mathbb{R}^2$, the median-level of a set of n lines has $O(n^{4/3})$ vertices [12]. Using efficient algorithms for constructing the median level of a set of lines (e.g., [7, 8, 18]), this dual transformation allows the projection median to be computed in $O(n^{4/3}\log^2 n)$ time in $\mathbb{R}^2$ [17], instead of $O(n^{2d-2}) = O(n^2)$ time. Using the algorithm of Agarwal and Matoušek [1] for computing k-levels in $\mathbb{R}^d$, Basu et al. [5] generalize this idea to higher dimensions and describe a dual transformation and a corresponding algorithm for computing the projection median in $O(n^{d(1-\frac{\delta_d}{d+1})+\epsilon})$ time, for some $\delta_d > 0$.

Both definitions of the projection median suggest that the number of components necessary for its computation is at least proportional to the number of times the median changes in the projection. As such, it appears unlikely that the projection median of a set of n points in $\mathbb{R}^d$ can be computed asymptotically faster than the time required to compute the median-level of a set of n hyperplanes in $\mathbb{R}^d$. This motivates the exploration of approximation algorithms that return a good approximation of the position of the projection median in significantly lower time. While approximating the projection median by a discretization such as a Riemann sum works well in $\mathbb{R}^2$, doing so is often computationally inefficient in $\mathbb{R}^d$. To obtain a desired overall closeness of approximation of $O(\varepsilon)$ requires using a $(d-1)$-dimensional grid of $\Theta(md)$ steps in each coordinate, where $m = \varepsilon^{-1}$, for a total of $\Theta((md)^{d-1})$ steps (see Davis and Rabinowitz [10, Section 5.5]). As the operations described in Algorithm 1 must be performed at each step, the resulting running time would be $\Omega((md)^{d-1}nd)$, making this approach unrealistic in most cases. In Sections 3.2 and 3.3 we describe and analyze two randomized approximation algorithms, both of which have running time $O(mnd)$ and, with high probability, achieve the desired closeness of approximation in $\mathbb{R}^d$ using only m randomly selected unit vectors.

3.2 Approximating the Projection Median in $O(mnd)$ Time

Taking advantage of representations (1) and (2), we propose two different approaches to approximate the projection median by Monte Carlo simulation. See Robert and Casella [25] for an excellent introduction to Monte Carlo simulation methods. Let U denote a randomly selected point on S^{d-1}. The point U uniquely identifies a randomly selected direction from the origin and l_U, the line passing through the origin and U. Then, starting with (1), we can write

$$\mathcal{M}_{\mathrm{P}}(P) = \frac{d}{A}\int_{S^{d-1}} \mathrm{med}(P_u)\,\mathrm{d}u = d\int_{S^{d-1}} \mathrm{med}(P_u) f_U(u)\,\mathrm{d}u = d\,\mathbf{E}[\mathrm{med}(P_U)], \tag{10}$$

where f_U is the uniform density over S^{d-1} and $\mathbf{E}$ denotes mathematical expectation. In other words, the above integral is the mathematical expectation of the median of the set of points P projected onto the line l_U, where U is selected uniformly at random on the unit sphere.

This suggests using, as a simple Monte Carlo approximation of $\mathcal{M}_{\mathrm{P}}(P)$, the sample average

$$\widehat{\mathcal{M}}_{\mathrm{P},1}(P) = \frac{d}{m}\sum_{j=1}^{m} \mathrm{med}(P_{U_j}), \tag{11}$$

where $U_1, U_2, \ldots, U_m$ are (pseudo) random unit vectors uniformly distributed over S^{d-1}. Calculating this approximation is relatively straightforward. First generate the m random unit vectors $U_1, U_2, \ldots, U_m$. This takes $O(md)$ time, independently of n. Then, calculate the median $\mathrm{med}(P_{U_j})$ of all the points in P projected onto l_{U_j}. This can be done in two steps for each index $j \in \{1, 2, \ldots, m\}$: first identify which projected point is associated with the median, which takes $O(nd)$ time, then calculate the coordinates of the projected median, which takes $O(d)$ time. Doing this for every point requires $O(mnd)$ time. Finally, consider the average of the medians obtained as an approximation of $\mathcal{M}_{\mathrm{P}}(P)$. As this last step takes $O(md)$ time, the resulting approximation is computed in $O(mnd)$ time. The following algorithm summarizes this approximation method.

Algorithm 1. *For $j = 1$ to m:*

1. *Generate U_j uniformly on S^{d-1}.*
2. *Find $p_*(U_j)$, the point whose projection is $\mathrm{med}(P_{U_j})$.*
3. *Calculate $\mathrm{med}(P_{U_j})$ as the projection of $p_*(U_j)$.*

Then, calculate the scaled coordinate-wise average of $\mathrm{med}(P_{U_1}), \mathrm{med}(P_{U_2}), \ldots, \mathrm{med}(P_{U_m})$ given in (11).

Note that Step 2 of Algorithm 1 can be completed without calculating the projections of all points in P. Specifically, it suffices to find the univariate median of the set of inner products $\langle p, U_j \rangle$ for all $p \in P$. Step 3 is then necessary to find the actual median projection onto l_{U_j}.

We now examine the probability of closeness of this approximation to the projection median $\mathcal{M}_{\mathrm{P}}(P)$. Specifically, this approximation *converges in probability* to the true projection median as $m \to \infty$, that is, for any set of points $P \subseteq \mathbb{R}^d$,

$$\forall \varepsilon > 0, \qquad \lim_{m\to\infty} \mathbf{Pr}(\|\widehat{\mathcal{M}}_{\mathrm{P},1}(P) - \mathcal{M}_{\mathrm{P}}(P)\|_2 < \varepsilon) = 1. \tag{12}$$

More importantly, this convergence holds for any $d \geq 1$. The following lemma establishes a lower bound on the probability of closeness of the approximation $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$ in terms of m, d, and the desired approximation factor $\varepsilon > 0$.

Lemma 4. *For any set of points $P = \{p_1, p_2, \ldots, p_n\} \subseteq \mathbb{R}^d$, any $\varepsilon > 0$, and any $m \geq d$,*

$$\mathbf{Pr}(\|\widehat{\mathcal{M}}_{\mathrm{P},1}(P) - \mathcal{M}_{\mathrm{P}}(P)\|_2 < \varepsilon) \geq 1 - C\, e^{-m\varepsilon^2/\delta_1(d,P)},$$

where $C = 1 + e^{5/12}/(\pi\sqrt{2})$, $\delta_1(d, P) = 8e^2 d^2 \|P\|^2$, and $\|P\| = \max_{1\leq i\leq n} \|p_i\|_2$.

Proof. We start by defining the random vectors $\mathrm{med}(P_{U_j}) = Y_j = (Y_{j1}, Y_{j2}, \ldots, Y_{jd})^{\mathsf{T}}$, where A^{T} denotes the transpose of A. Thus,

$$\widehat{\mathcal{M}}_{\mathrm{P},1}(P) = \frac{d}{m} \sum_{j=1}^{m} Y_j = d\bar{Y} = d(\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_d)^{\mathsf{T}},$$

where $\bar{Y}_k = \frac{1}{m} \sum_{j=1}^{m} Y_{jk}$ for $k = 1, 2, \ldots, d$. We note that, from (10), we have $\mathbf{E}[\bar{Y}_k] = \mu_k = \mathbf{E}[\mathrm{med}(P_U)]_k$, the kth component of $\mathcal{M}_{\mathrm{P}}(P)/d$. We can further define the random vectors $Z_j = Y_j - \mathbf{E}[\mathrm{med}(P_U)]$, which are independent and identically distributed with mean zero, and that all satisfy

$$\|Z_j\|_2 \leq \|Y_j\|_2 + \|\mathbf{E}[\mathrm{med}(P_U)]\|_2 \leq 2 \max_{1\leq i\leq n} \|p_i\|_2 = 2\|P\| \ . \tag{13}$$

Note that

$$d\bar{Z} = d(\bar{Y} - \mathbf{E}[\mathrm{med}(P_U)]) = \widehat{\mathcal{M}}_{\mathrm{P},1}(P) - \mathcal{M}_{\mathrm{P}}(P),$$

and a theorem of Prokhorov [24] allows us to write, for all $m \geq d$,

$$\mathbf{Pr}\left(\|\sqrt{m}\bar{Z}\|_2 \geq \sqrt{m}\varepsilon/d\right) \leq \left[1 + \frac{e^{5/12}}{\pi\sqrt{2}} \frac{\mathbf{E}[\|Z_1\|_2^2]}{4\|P\|^2}\right] e^{-m\varepsilon^2/\delta_1(d,P)} \leq Ce^{-m\varepsilon^2/\delta_1(d,P)},$$

since $\mathbf{E}[\|Z_1\|_2^2] \leq 4\|P\|^2$ trivially follows from (13). Using this, we get

$$\mathbf{Pr}(\|\widehat{\mathcal{M}}_{\mathrm{P},1}(P) - \mathcal{M}_{\mathrm{P}}(P)\|_2 < \varepsilon) = 1 - \mathbf{Pr}\left(\|\sqrt{m}\bar{Z}\|_2 \geq \sqrt{m}\varepsilon/d\right) \geq 1 - Ce^{-m\varepsilon^2/\delta_1(d,P)},$$

thus completing the proof. □

The preceding results highlight the contribution of the dimension d to the difficulty of approximating the projection median using (11). The value of d affects the quality of that approximation in the sense that large values of d lead to a smaller probability of closeness. Also, as will be seen below, the size of confidence regions on the projection median grows faster than d^2. The deterioration of the rate of convergence of the probability of closeness to 1, when d increases, comes from the use of the scaling factor d in the approximation to the projection median. It is not an artifact of the approach taken to derive the result, as Prokhorov's bound is the best possible in terms of d, only adding the requirement that $m \geq d$.

We also note that the bound given in the previous lemma can theoretically be improved by somehow centering the points. Indeed, if one were to use the point o as the origin, the previous result would still be valid but with $\|P\| = \max_{1\leq i\leq n} \|p_i - o\|_2$. The best choice for o would then be the Euclidean centre of P, i.e.,

$$o_* = \arg\min_{x\in\mathbb{R}^d} \max_{1\leq i\leq n} \|p_i - x\|_2.$$

Computing this best origin o_* can be done in $O(d^{O(d)}n)$ time [2], which may be prohibitive in high dimensions. Using this best origin could nevertheless lead to improved overall efficiency when n is very large but d is small, especially when considering multisets P that are highly concentrated around o_* and for which $\|o_*\|_2$ is large.

Lemma 4 makes it possible to derive a value of m which guarantees a certain probability of closeness for a desired approximation factor. Indeed, a desired probability of closeness p_c will be attained if

$$m \geq \frac{8e^2d^2}{\epsilon^2}\left[\log C - \log(1-p_c)\right],$$

where ε is the desired approximation factor and $\epsilon = \varepsilon/\|P\|$ is the tolerated relative error. In particular, we see that taking $m \in \Theta(\epsilon^{-2}d^2\log n)$ is sufficient to guarantee that our approximation is asymptotically close to the true projection median with high probability, i.e., to guarantee a probability of closeness $p_c \geq 1-n^{-a}$ asymptotically for some fixed $a \geq 1$. This gives the following theorem.

Theorem 5. *Choosing $m \in \Theta(\epsilon^{-2} d^2 \log n)$ guarantees that Algorithm 1 returns a point within relative distance ϵ of the true projection median with high probability. Also, this point is returned in time $\Theta(\epsilon^{-2} d^3 n \log n)$.*

Unfortunately, this result does not quite allow us to clearly assess the precision of the resulting approximation and, in most cases, is quite conservative: it suggests the use of unnecessarily large simulation sizes when p_c is close to 1. However, we can take advantage of another important aspect of Monte Carlo simulation. Specifically, when approximating $\mathcal{M}_{\mathrm{P}}(P)$ through Monte Carlo simulation, we can easily assess the precision of the approximation by constructing a confidence region for it. Asymptotically in m, we have

$$\widehat{\mathcal{M}}_{\mathrm{P},1}(P) \approx N_d\left(\mathcal{M}_{\mathrm{P}}(P), \frac{d^2}{m} V_1(P)\right), \tag{14}$$

where $N_d(\cdot,\cdot)$ denotes the d-multivariate normal distribution, and the $d \times d$ scatter matrix $V_1(P)$ is constructed from the simulated data using

$$V_1(P) = \frac{1}{m-1} \sum_{j=1}^{m} (Y_j - \bar{Y})(Y_j - \bar{Y})^{\mathsf{T}},$$

with the notation introduced in the proof of Lemma 4. This is simply the $d \times d$ sample covariance matrix constructed from the medians obtained from the random projections onto $l_{U_1}, l_{U_2}, \ldots, l_{U_m}$. Hence, $V_1(P)$ is calculated only once at the end of the Monte Carlo experiment using simple matrix operations. Now, the approximate normality of $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$ given by (14) allows us to estimate its precision through the construction of a confidence ellipsoid using the fact that

$$\lim_{m\to\infty} \mathbf{Pr}\Big(md^{-2}[\widehat{\mathcal{M}}_{\mathrm{P},1}(P) - \mathcal{M}_{\mathrm{P}}(P)]^{\mathsf{T}} V_1(P)^{-1} [\widehat{\mathcal{M}}_{\mathrm{P},1}(P) - \mathcal{M}_{\mathrm{P}}(P)] \le \varepsilon\Big) = H_d(\varepsilon), \tag{15}$$

where the function H_d is the cumulative distribution function of the chi-square distribution with d degrees of freedom. In practice, we can fix an acceptable confidence level $1-\alpha$ for some $\alpha \in (0,1)$; $\alpha = 0.05$ is often reasonable. Then, we find ε such that $H_d(\varepsilon) = 1-\alpha$, or

$$\varepsilon = H_d^{-1}(1-\alpha), \tag{16}$$

known as the quantile of order $1-\alpha$ of the chi-square distribution with d degrees of freedom. According to (15), our Monte Carlo experiment has a probability tending to $1-\alpha$ of leading to a hyperellipsoid containing $\mathcal{M}_{\mathrm{P}}(P)$ as the size of the simulation $m \to \infty$. Note that ε given in (16) is a strictly increasing (and unbounded) function of d. Hence, keeping everything else constant, the confidence region given by

$$[x - \widehat{\mathcal{M}}_{\mathrm{P},1}(P)]^{\mathsf{T}} V_1(P)^{-1} [x - \widehat{\mathcal{M}}_{\mathrm{P},1}(P)] \le \frac{d^2}{m} H_d^{-1}(1-\alpha), \tag{17}$$

and corresponding to the interior and the boundary of the hyperellipsoid, rapidly increases in size with increasing d.

Alternatively, we can construct intervals that have probability at least $1-\alpha$ of simultaneously containing each of the d individual components of the projection median. By the Bonferroni inequality, we have

$$\lim_{m\to\infty} \mathbf{Pr}\Big(\bigcap_{k=1}^{d}\Big\{\frac{\sqrt{m}}{d\sqrt{v_{kk}}}|\widehat{\mathcal{M}}_{\mathrm{P},1}(P)_k-\mathcal{M}_{\mathrm{P}}(P)_k|\leq\varepsilon\Big\}\Big)\geq 1-d[1-\Phi(\varepsilon)], \tag{18}$$

where Φ is the cumulative distribution function of the (univariate) standard normal distribution and v_{kk} denotes the diagonal elements of $V_1(P)$. Here, it is common to use $\varepsilon=\Phi^{-1}(1-\alpha/d)$, the quantile of order $1-\alpha/d$. The lower bound on the right-hand side of (18) becomes $1-\alpha$, the same as in the previous case. This bound is known to be quite conservative in many cases, especially so when d is large. The interpretation of (18) is otherwise the same as that of (15): we are conducting a Monte Carlo experiment that has a probability of at least $1-\alpha$ to lead to a collection of d individual intervals that each contain their associated component of $\mathcal{M}_{\mathrm{P}}(P)$.

3.3 Approximating the Projection Median Using a Weighted Mean

Another approximation to the projection median relies on the weighted average representation given in (2) and the fact that the weights w_p can be rewritten as

$$w_p=\frac{\int_{S_p^{d-1}}1\ \mathrm{d}u}{\int_{S^{d-1}}1\ \mathrm{d}u}=\frac{1}{A}\int_{S_p^{d-1}}1\ \mathrm{d}u=\int_{S_p^{d-1}}f_U(u)\ \mathrm{d}u=\mathbf{Pr}(U\in S_p^{d-1})=\mathbf{Pr}(p_*(U)=p),$$

where S_p^{d-1} is defined in (3) and U is selected uniformly at random over the unit sphere. That is, w_p corresponds to the proportion of points $u\in S^{d-1}$ such that the median of P_u is associated with p. This suggests another Monte Carlo scheme for approximating the projection median $\mathcal{M}_{\mathrm{P}}(P)$. We can calculate

$$\widehat{\mathcal{M}}_{\mathrm{P},2}(P)=\sum_{p\in P}\widehat{W}_p p,\quad \text{where}\quad \widehat{W}_p=\frac{1}{m}\sum_{j=1}^{m}\mathbf{I}_{S_p^{d-1}}(U_j), \tag{19}$$

and $\mathbf{I}_A(U)=1$ if $U\in A$, 0 otherwise, so that $\widehat{W}_p$ corresponds to the sample proportion of $U_1,U_2,\ldots,U_m$ that belong to S_p^{d-1}. Note that, trivially, $\sum_{p\in P}\widehat{W}_p=1$. This approximation is easier and faster to calculate than $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$, although the number of operations is asymptotically of the same order. Again, start by generating the m random unit vectors $U_1,U_2,\ldots,U_m$ in $O(md)$ time. Then, for each line l_{U_j}, identify which projected point of P is associated with the median of P_U, which takes $O(nd)$ time. Note that here, the projected median does not have to be calculated, as the original points are now used instead of the projected points. Doing this for every point then requires $O(mnd)$ time. Finally, consider the weighted average of the original points in P as an approximation of $\mathcal{M}_{\mathrm{P}}(P)$. As this last step takes $O(md)$ time, the resulting approximation is computed in $O(mnd)$ time. The following algorithm summarizes this second approximation method.

Algorithm 2. *For $j = 1$ to m:*

1. *Generate U_j uniformly on S^{d-1}.*
2. *Find $p_*(U_j)$, the point p such that $\mathbf{I}_{S_p^{d-1}}(U_j) = 1$.*

Then, calculate the weight $\widehat{W}_p$ associated to each point $p \in P$ and the weighted average (19).

As before, Step 2 of Algorithm 2 can be completed without calculating the projections of all points in P. The following result gives a concentration inequality for $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$. The approximation to $\mathcal{M}_{\mathrm{P}}(P)$ provided by Algorithm 2 does not seem to deteriorate nearly as fast as that of $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$ as the dimension d increases. Note that $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$ converges in probability to the true projection median, i.e. it also satisfies (12).

We now derive two bounds that allow us to characterize the precision of Algorithm 2. The first bound is a straightforward adaptation of Lemma 4. The second is based on a multinomial inequality due to Devroye and Györfi [11].

Lemma 6. *For any set of points $P \subseteq \mathbb{R}^d$, any $\varepsilon > 0$, and any $m \geq d$,*

$$\mathbf{Pr}(\|\widehat{\mathcal{M}}_{\mathrm{P},2}(P) - \mathcal{M}_{\mathrm{P}}(P)\|_2 < \varepsilon) \geq 1 - C\, e^{-m\varepsilon^2/\delta_2(P)},$$

where $C = 1 + e^{5/12}/(\pi\sqrt{2})$ and $\delta_2(P) = 8e^2\|P\|^2$.

Proof. For $j = 1, \ldots, m$, define the d-dimensional random vectors $Y_j = p_*(U_j)$ with expectation

$$\mathbf{E}[Y_j] = \sum_{p\in P} p\, \mathbf{Pr}(p_*(U) = p) = \mathcal{M}_{\mathrm{P}}(P).$$

Proceeding as in the proof of Lemma 4, let $Z_j = Y_j - \mathbf{E}[Y_j]$ and note, once again, that $\|Z_j\|_2 \leq 2\|P\|$. Then, we have that $\bar{Z} = \widehat{\mathcal{M}}_{\mathrm{P},2}(P) - \mathcal{M}_{\mathrm{P}}(P)$ and as before,

$$\mathbf{Pr}\left(\left\|\widehat{\mathcal{M}}_{\mathrm{P},2}(P) - \mathcal{M}_{\mathrm{P}}(P)\right\|_2 < \varepsilon\right) = 1 - \mathbf{Pr}\left(\|\sqrt{m}\bar{Z}\|_2 \geq \sqrt{m}\varepsilon\right) \geq 1 - Ce^{-m\varepsilon^2/\delta_1(1,P)}.$$

The conclusion follows from the fact that $\delta_2(P) = \delta_1(1, P)$. □

As before, this result allows us to see that a desired probability of closeness p_c will be attained if

$$m \geq \max\left(\frac{8e^2}{\epsilon^2}\left[\log C - \log(1 - p_c)\right], d\right), \tag{20}$$

where $\epsilon = \varepsilon/\|P\|$ is the tolerated relative error. As observed by Prokhorov [24], d has little effect on the choice of m since $\epsilon^{-2}\log n$ will typically be much larger than d. In this case, taking $m \in \Theta(\epsilon^{-2}\log n)$ is sufficient to guarantee that our approximation is asymptotically close to the true projection median with high probability, i.e., to guarantee a probability of closeness $p_c \geq 1 - n^{-a}$ asymptotically for some fixed $a \geq 1$.

When d is very large, choosing $m \in \Theta(d)$ guarantees that the approximation is asymptotically close to the true projection median with high probability. Interestingly, it is also possible to derive a bound similar to the ones obtained so far by linking the probability of closeness (for the projection median) to the probability that a multinomial random vector is close to its mean. This gives another condition on m, this one completely independent of d, that also guarantees a desired probability of closeness. This is an important property of our second approximation: it does not require increasingly large simulation sizes as d increases, i.e., the choice of m is independent of d when d is large.

Lemma 7. *For any set of points $P \subseteq \mathbb{R}^d$ and any $\varepsilon > 0$, if m is sufficiently large (i.e., $m \geq 20\|P\|^2 n/\varepsilon^2$), then*

$$\mathbf{Pr}\left(\|\widehat{\mathcal{M}}_{\mathrm{P},2}(P) - \mathcal{M}_{\mathrm{P}}(P)\|_2 < \varepsilon\right) \geq 1 - 3e^{-m\varepsilon^2/\delta_3(P)},$$

where $\delta_3(P) = 25\|P\|^2$.

Proof. First, we observe that

$$\begin{aligned}\mathbf{Pr}\left(\|\widehat{\mathcal{M}}_{\mathrm{P},2}(P) - \mathcal{M}_{\mathrm{P}}(P)\|_2 \geq \varepsilon\right) &= \mathbf{Pr}\left(\left\|\sum_{i=1}^{n}(\widehat{W}_{p_i} - w_{p_i})p_i\right\|_2 \geq \varepsilon\right)\\ &\leq \mathbf{Pr}\left(\|P\|\sum_{i=1}^{n}|\widehat{W}_{p_i} - w_{p_i}| \geq \varepsilon\right)\end{aligned}$$

by the triangle inequality. The result then follows from Lemma 3.1 of Devroye and Györfi [11] since $m(\widehat{W}_{p_1}, \widehat{W}_{p_2}, \ldots, \widehat{W}_{p_n})'$ has a multinomial distribution with parameters m and $w_{p_1}, w_{p_2}, \ldots, w_{p_n}$. The inequality is valid for all m such that $n/m \leq (\varepsilon/\|P\|)^2/20$. ☐

As mentioned above, this bound does not depend on d other than indirectly through $\|P\|$. This suggests that the performance of $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$ might deteriorate very slowly as d increases, but that it would do so at a much slower rate than $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$. Indeed, the inequality given in Lemma 7 is directly comparable to the ones obtained in Lemmas 4 and 6, and it suggests a significant difference in the rate of convergence of the probability of closeness to 1 as d increases. The difference is also highlighted by the fact that, for a given approximation factor ε, a desired probability of closeness p_c will be attained if

$$m \geq \frac{5}{\epsilon^2}\max\left\{5\left[\log(3) - \log(1-p_c)\right], 4n\right\}, \tag{21}$$

where $\epsilon = \varepsilon/\|P\|$ is the tolerated relative error. In particular, we see that taking $m \in \Theta(\epsilon^{-2}n)$ is sufficient to guarantee that our approximation is asymptotically close to the true projection median with high probability, i.e., to guarantee a probability of closeness $p_c \geq 1 - n^{-a}$ asymptotically for some fixed $a \geq 1$. Finally, combining (20) and (21), we get the following theorem, which establishes the performance of $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$.

Theorem 8. *Choosing $m \in \Theta\big(\min(d+\epsilon^{-2}\log n, \epsilon^{-2}n)\big)$ guarantees that Algorithm 2 returns a point within relative distance ϵ of the true projection median with high probability. Also, this point is returned in time $\Theta\big(dn \cdot \min(d+\epsilon^{-2}\log n, \epsilon^{-2}n)\big)$.*

Interestingly, for large set size n and small d, this simplifies to the result of Theorem 5. However, for small n and large d, it reduces to using $m \in \Theta(\epsilon^{-2}n)$ and a total

Table 1: Computation time for approximating the true projection median using the two approximation methods in various scenarios. In each case, the relative distance ϵ is held constant and $0 < \alpha < 1 < \beta$.

scenario	$\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$	$\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$
d held constant, $n \to \infty$	$\Theta(n \log n)$	$\Theta(n \log n)$
$d \to \infty$, n held constant	$\Theta(d^3)$	$\Theta(d)$
$n \to \infty$, $d \in \Theta\big((\log n)^\alpha\big)$	$\Theta\big(n(\log n)^{1+3\alpha}\big)$	$\Theta\big(n(\log n)^{1+\alpha}\big)$
$n \to \infty$, $d \in \Theta\big((\log n)^\beta\big)$	$\Theta\big(n(\log n)^{1+3\beta}\big)$	$\Theta\big(n(\log n)^{2\beta}\big)$
$n \to \infty$, $d \in \Theta(n^\alpha)$	$\Theta(n^{1+3\alpha} \log n)$	$\Theta\big(n^{1+2\alpha}\big)$
$n \to \infty$, $d \in \Theta(n^\beta)$	$\Theta(n^{1+3\beta} \log n)$	$\Theta\big(n^{2+\beta}\big)$

time of $\Theta(\epsilon^{-2} d n^2)$. The result compares extremely well to the computing time obtained in Theorem 5, which increases as d^3 when other parameters are kept constant. Given that $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$ is also easier to calculate, it seems clear that this second approximation will be preferable in general. We summarize the implications of Theorem 8 in Table 1.

Now, we can assess the precision of $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$ by constructing a confidence region for it. Asymptotically in m, we have that $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$ is approximately distributed as $N_d(\mathcal{M}_{\mathrm{P}}(P), \frac{1}{m} V_2(P))$, where the $d \times d$ scatter matrix $V_2(P)$ is constructed from the simulated data using

$$V_2(P) = \frac{m}{m-1} \sum_{i=1}^{n} \widehat{W}_{p_i} \left(p_i - \widehat{\mathcal{M}}_{\mathrm{P},2}(P)\right) \left(p_i - \widehat{\mathcal{M}}_{\mathrm{P},2}(P)\right)^{\mathsf{T}}.$$

Following the same approach as before, the confidence ellipsoid with asymptotic probability $1 - \alpha$ of containing the projection median is given by

$$[x - \widehat{\mathcal{M}}_{\mathrm{P},2}(P)]^{\mathsf{T}} V_2(P)^{-1} [x - \widehat{\mathcal{M}}_{\mathrm{P},2}(P)] \leq \frac{1}{m} H_d^{-1}(1 - \alpha). \tag{22}$$

The size of this hyperellipsoid increases slowly with d. It is also possible to construct simultaneous confidence intervals for the components of the projection median by essentially adapting (18) to the current context.

4 Simulation and Analysis

In this section, we briefly describe a simulated experiment we ran to examine empirically the precision of the two approximation algorithms described in Sections 3.2 and 3.3,

The first scenario considered is where P is a set of $n = 9$ points in $\mathbb{R}^2$. These points are given in Table 2. We repeated the Monte Carlo experiment three times each with $m = 50$ and $m = 150$ randomly generated angles. For each simulation we constructed the

Table 2: The points in $\mathbb{R}^2$ used for the first and second simulation experiments, and the set of weights $w(P)$ and $w(P')$ associated to each scenario.

x_k	y_k	$w(P)$	$w(P')$
0.6276	0.2772	0.0358	–
10.0	10.0	–	0.0008
0.2552	0.5599	0.0137	0.0137
0.1545	0.0257	0.0716	0.0250
0.8286	0.0545	0.1159	0.1159
0.4089	0.8235	0.0133	0.0133
0.5792	0.0441	0.0234	0.0234
0.2365	0.1227	0.1343	0.2225
0.4967	0.3219	0.4904	0.4837
0.6811	0.1024	0.1016	0.1016

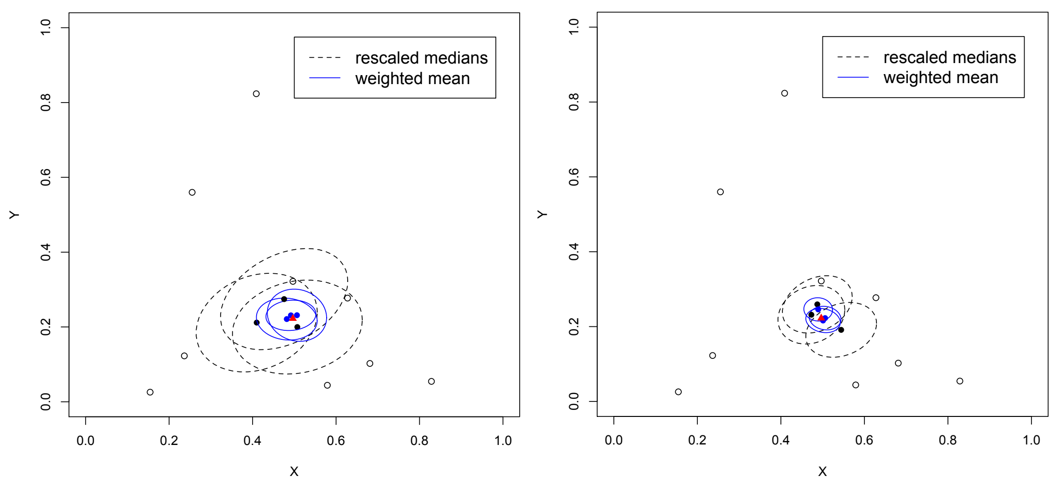


Figure 2: Confidence ellipses obtained from three repetitions of the Monte Carlo experiment with $m = 50$ random angles (left) and $m = 150$ (right). The projection median is shown in red.

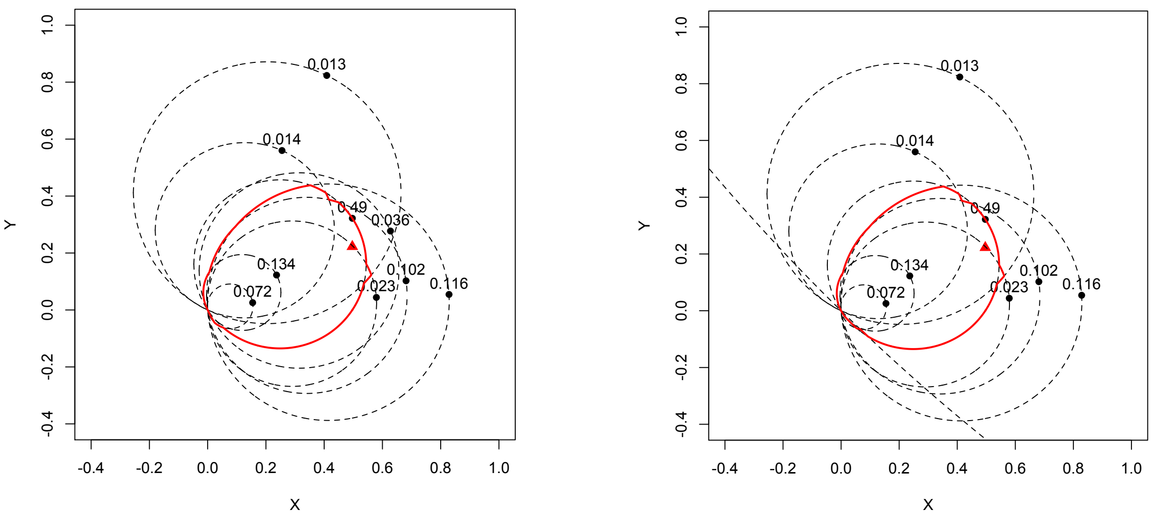


Figure 3: Sets of points P (left) and P' (right) displayed with the trajectory of the projection of each point (dashed black circle), the trajectory of the corresponding projected medians (solid red arcs), the weight associated with each point, and the projection median (larger red point). The point $p'_1 = (10, 10)$ with weight 0.0008 is not displayed in the plot of P'.

associated confidence ellipses (17) and (22) using a confidence level of 0.95. From Figure 2, it seems clear that $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$ leads to smaller confidence regions than $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$ does. The average area contained in the confidence ellipses shown in Figure 2 is 0.0110 and 0.0037 based on $\widehat{\mathcal{M}}_{\mathrm{P},2}(P)$, compared to 0.0552 and 0.0187 based on $\widehat{\mathcal{M}}_{\mathrm{P},1}(P)$, for simulations of size 50 and 150, respectively. Hence, in the current context, the second method is about five times more precise than the first one.

Now, to test the robustness of the projection median to the presence of outliers, we briefly consider the situation where the first point is replaced by the point $p'_1 = (10, 10)$. The projection medians are given by $\mathcal{M}_{\mathrm{P}}(P) = (0.4966, 0.2225)$ and $\mathcal{M}_{\mathrm{P}}(P') = (0.4928, 0.2284)$, showing the robustness of the projection median to the presence of outliers. Observe that the weights of only a few points change when p_1 changes. The weight of p_1 goes from 0.0358 down to 0.0008, while the weights of p_2, p_4, p_5, p_6 and p_9 are unaffected (see Table 2). This is seen in Figure 3, where removing p_1 and replacing it with p'_1 affects the median level of the trajectories obtained from projecting each point.

5 Geometric Properties of the Projection Median

In Section 5.1 we review known geometric properties of the projection median and apply its definition as a weighted mean to extend these and to establish new properties relevant to estimating location of geometric multivariate data. In Section 5.2 we compare the projection median against other high-dimensional generalizations of the median in terms of these properties.

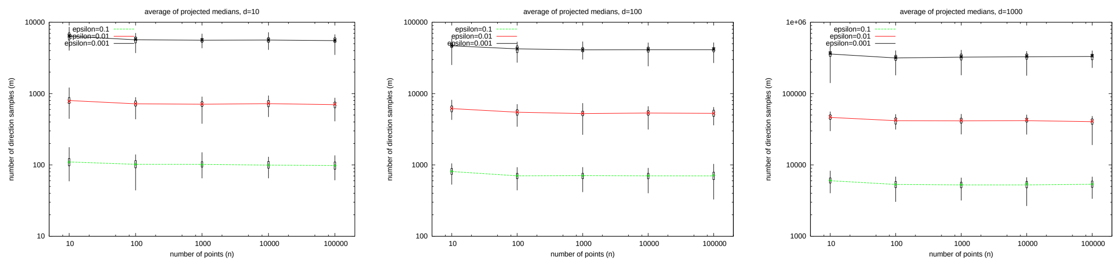


Figure 4:	algorithm	Algorithm 1 (average of projected medians)
	number of data	$n \in \{10^1, 10^2, 10^3, 10^4, 10^5\}$
	error tolerance	$\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}\}$
	data dimensionality	$d = 10^1$ (left), $d = 10^2$ (centre), $d = 10^3$ (right)

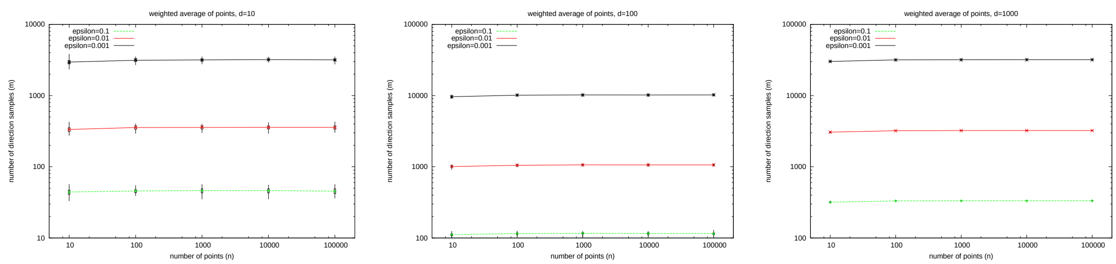
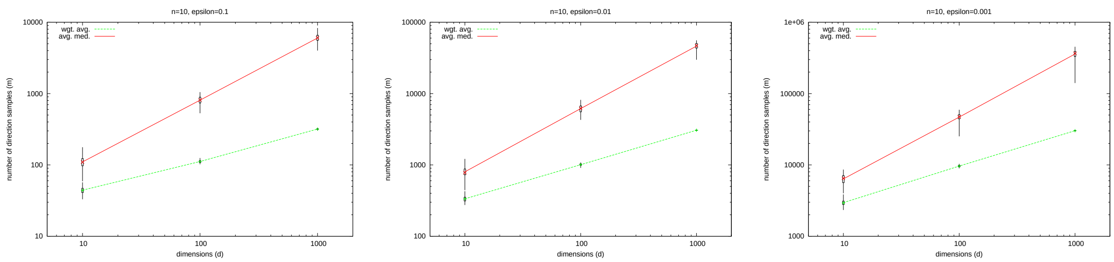


Figure 5:	algorithm	Algorithm 2 (weighted mean)
	number of data	$n \in \{10^1, 10^2, 10^3, 10^4, 10^5\}$
	error tolerance	$\epsilon \in \{10^{-3}, 10^{-2}, 10^{-1}\}$
	data dimensionality	$d = 10^1$ (left), $d = 10^2$ (centre), $d = 10^3$ (right)

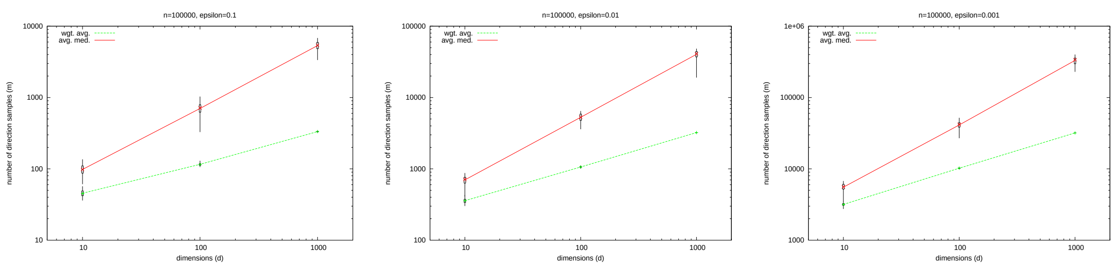
5.1 Properties of the Projection Median

The projection median was shown to provide a stable approximation of the Weber point that can be computed exactly. Specifically, Durocher and Kirkpatrick [17] showed that the sum of the distances from the projection median to points in $P \subseteq \mathbb{R}^2$ is at most $4/\pi$ times that of the corresponding sum for the Weber point, and that any ϵ-perturbation of points in P results in a displacement of at most $4\epsilon/\pi$ in the projection median of P (unlike previous approximations that are not κ-stable for any fixed κ). Durocher [14] extended the latter result to show that in $\mathbb{R}^3$, any ϵ-perturbation of points in P results in a displacement of at most $3/2$ in the projection median of P. Basu et al. [5] generalized these results to higher dimensions, showing that in $\mathbb{R}^d$ the sum of the distances from the projection median to points in P is at most $(d/\pi)B(d/2, 1/2)$ times that of the corresponding sum for the Weber point and that any ϵ-perturbation of points in P results in a displacement of at most $(d/\pi)B(d/2, 1/2)$ in the projection median of P, where $B(\alpha, \beta)$ denotes the beta function. The following upper bounds hold:

$$\forall d \geq 2,\ \frac{d}{\pi}B(d/2, 1/2) \leq \frac{2\sqrt{2d}}{\pi} < \sqrt{d} \qquad \text{and} \qquad \lim_{d\to\infty} \frac{d}{\pi}B(d/2, 1/2) = \sqrt{\frac{2d}{\pi}}\ . \tag{23}$$



	algorithms	Algorithm 1 (average of projected medians) and Algorithm 2 (weighted mean)
Figure 6:	number of data	$n = 10^1$
	error tolerance	$\epsilon = 10^{-1}$ (left), $\epsilon = 10^{-2}$ (centre), $\epsilon = 10^{-3}$ (right)
	data dimensionality	$d \in \{10^1, 10^2, 10^3\}$



	algorithms	Algorithm 1 (average of projected medians) and Algorithm 2 (weighted mean)
Figure 7:	number of data	$n = 10^5$
	error tolerance	$\epsilon = 10^{-1}$ (left), $\epsilon = 10^{-2}$ (centre), $\epsilon = 10^{-3}$ (right)
	data dimensionality	$d \in \{10^1, 10^2, 10^3\}$

A similarity transformation is an affine transformation that preserves not only collinearity and relative distances, but also angles. The projection median was shown to be equivariant under similarity transformations in two and three dimensions [14]. While the earlier proofs can be generalized to higher dimensions, the generalization is straightforward using the weighted mean definition of the projection median. Given any multiset of points P in $\mathbb{R}^d$, any unit vector $u \in S^{d-1}$, any similarity transformation $f : \mathbb{R}^d \to \mathbb{R}^d$, and any point $p \in P$, it follows that p_u is a median of P_u if and only if $f(p)_{f(u)}$ is a median of $f(P)_{f(u)}$, where q_v denotes the projection of the point $q \in \mathbb{R}^d$ onto the line l_v for any given $v \in S^{d-1}$. Consequently, for any $p \in P$, the weight w_p for p with respect to P is equal to the weight $w_{f(p)}$ for $f(p)$ with respect to $f(P)$. Therefore, $f(\mathcal{M}_{\mathrm{A}}(P)) = \mathcal{M}_{\mathrm{A}}(f(P))$, giving the following corollary of Theorem 2.

Corollary 9. *The projection median is equivariant under similarity transformations.*

Basu et al. [5] showed that the projection median has a breakdown point of $\alpha = 1/2$. That is, if fewer than αn points are perturbed in any set P of n points, then $\mathcal{M}_{\mathrm{P}}(P)$ remains within the relative neighbourhood of the unperturbed points, where the perturbed points can be moved any arbitrary distance.

The original definition of the projection median was also shown to be consistent from two to three dimensions [14]. That is, the three-dimensional projection median of a set of co-planar points in $\mathbb{R}^3$ coincides with the two-dimensional projection median of the corresponding set of points in $\mathbb{R}^2$. Choose any $d \geq 1$ and any set P of points in $\mathbb{R}^d$ that lies in some $(d-1)$-dimensional flat H of $\mathbb{R}^d$. For any unit vector $u \in S^{d-1}$ and any point $p \in P$, p_u is a median of P_u if and only if p_{u_H} is a median of P_{u_H}, where u_H denotes the normalized projection of u onto H. Consequently, the d-dimensional weight of p in $\mathbb{R}^d$ is equal to its $(d-1)$-dimensional weight in H. Consistency from dimension d to $d+1$ follows immediately from the weighted mean definition of the projection median, giving the following corollary of Theorem 2.

Corollary 10. *The definition of the projection median is consistent across dimensions.*

The weights in the weighted sum definition of the projection median are non-negative and sum to 1. Thus, the projection median of P is a convex combination of the elements of P. This immediately implies that $\mathcal{M}_{\mathrm{P}}(P)$ is local to P (since $\mathcal{M}_{\mathrm{P}}(P)$ is contained within the convex hull of P) and that $\mathcal{M}_{\mathrm{P}}(P)$ is unique.

5.2 Comparison Against Other Multivariate Estimators

The univariate median can be generalized to higher dimensions in many ways. The Weber point and the projection median are two such definitions of high-dimensional medians central to this work. Extending a similar table compiled by Basu et al. [5], in Table 3 we summarize relevant properties of the projection median and compare these against the Weber point, as well as against two other common multivariate estimators of location: the rectilinear median and the centre of mass. These properties include the sum of the distances to points in P relative to the minimum (recall that the optimum is attained by the Weber point); instability (see Section 5); breakdown point, i.e., the fraction α of P that must be displaced before the estimator moves away from the unperturbed points of P; whether the estimator can be expressed as a weighted mean; invariance under similarity transformations (rotation, reflection, uniform scaling, translation); consistency across dimensions, i.e., whether the estimator of P is equal under its k- and d-dimensional definitions when P lies in a k-flat of $\mathbb{R}^d$ for any $k \leq d$; whether the estimator is uniquely defined; the running time of the fastest known algorithm for computing the exact position of the estimator; and the running time of the fastest known algorithm for computing a good approximation of the exact position of the estimator (assuming a sufficiently small fixed approximation factor ϵ).

A rectilinear median (also known as coordinatewise median) is given by finding a univariate median independently for each coordinate. As such, a rectilinear median of a set of points P minimizes the sum of rectilinear (ℓ_1) distances to points in P. The centre of mass is given by the coordinatewise weighted mean of P, where each point in a multiset P is weighted by its multiplicity. The centre of mass is easily shown to be the unique point that minimizes the sum of the squared distances to points in P [26, 30].

Table 3: A comparison of properties of common multivariate estimators of location

Property	Weber point	Rectilinear median	Centre of mass	Projection median
relative sum of distances	1	$\sqrt{d}$ [14]	2 [14]	$(d/\pi)B(d/2,1/2)$ [5][2]
instability	∞ [17]	$\sqrt{d}$ [14]	1	$(d/\pi)B(d/2,1/2)$ [5][2]
breakdown point	$1/2$	$1/2$	$1/n$	$1/2$
weighted mean	✗	✗	✓	✓
similarity transformations	✓	✗	✓	✓
dimensional consistency	✓	✗	✓	✓
uniqueness	✓[3]	✓[4]	✓	✓
exact computation	✗	$O(nd)$	$O(nd)$	$O(n^{d+\epsilon})$
approximate computation	$O(nd)$ [20, 6]	$O(nd)$	$O(nd)$	$O(\min\{n^2d, nd^3 \log n\})$

In addition to the four listed in Table 3, many other estimators of multivariate location exist, including various medians (deepest points) defined according to data depth measures, e.g., simplicial median, Tukey (half-space) median (which generalizes centre points), Oja median, majority median, convex hull peeling median, zonoid median, Mahalanobis median, Gao spatial rank median, etc. Most of these are unstable or do not guarantee any bounded approximation of the relative sum of distances. See Aloupis [3], Small [28], and Zuo and Serfling [32] for surveys of data depth.

6 Discussion and Directions for Additional Research

In this paper we introduced a new definition of the projection median as a weighted mean and show its equivalence to the original definition. The form of this new definition is more consistent with those of traditional statistical estimators of location, allows many of the geometric properties of the projection median to be established more easily, and suggests randomized algorithms that compute approximations of the projection median with increased accuracy and efficiency.

Several natural questions remain unanswered. With respect to approximation algorithms, it is unknown how closely the projection median can be approximated in $O(f(\epsilon)nd)$ time, for some function f. The Weber point can be approximated in $O(f(\epsilon)nd)$ time [6, 9, 20]; our current approximation algorithms for the projection median achieve $O(g(\epsilon)n^2d)$ and $O(g(\epsilon)nd^3 \log n)$ time, for $g(\epsilon) = \epsilon^{-2}$. Can the best of both worlds be achieved simultaneously? For example, does there exist a fixed $c > 0$ such that for any multiset of n points $P \subseteq \mathbb{R}^d$ and any $\epsilon > 0$, an ϵ-approximation of the projection median of P can be returned in $O(\epsilon^{-c}nd)$ time? If so, what is the smallest such c?

[2] As shown in (23), for all $d \geq 2$, $(d/\pi)B(d/2, 1/2) \leq 2\sqrt{2d}/\pi < \sqrt{d}$, where $B()$ denotes the beta function.
[3] If $|P|$ is odd or there exist three points in P that are not collinear, then the Weber point is unique.
[4] If $|P|$ is odd, then the rectilinear median is unique.

As mentioned in Section 1, if the minimum (or maximum) of P_u is used in (1) instead of $\text{med}(P_u)$, the result is the Steiner centre of P, which, as shown by Shephard [27], can also be expressed as a weighted mean. For the Steiner centre, the weight, w_p, of each point $p \in P$ corresponds to the exterior solid turn angle at p on the convex hull of P (interior points have weight 0). Upon normalizing, this weight has an interpretation analogous to (2), corresponding to the proportion of unit vectors $u \in S^{d-1}$ such that the projection of p onto l_u is the leftmost point of P_u relative to u. This leads to a natural conjecture that the equivalence holds for any order statistic, which we show to be true in Corollary 3. As mentioned in Section 2.3, Corollary 3 applies to multivariate trimmed means. It could be interesting to identify applications for this generalization, and to compare it against other definitions of trimmed means in higher dimensions [13, 22, 31].

References

[1] P. K. Agarwal and J. Matoušek. Dynamic half-space range reporting and its applications. *Algorithmica*, 13(4):325–345, 1995.

[2] P. K. Agarwal, M. Sharir, and S. Toledo. An efficient multi-dimensional searching technique and its applications. Technical Report CS-1993-20, Duke University, 1993.

[3] G. Aloupis. Geometric measures of data depth. In R. Liu, R. Serfling, and D. Souvaine, editors, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 72 Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, pages 147–158, 2006.

[4] C. Bajaj. The algebraic degree of geometric optimization problems. *Discrete and Computational Geometry*, 3:177–191, 1988.

[5] R. Basu, B. B. Bhattacharya, and T. Talukdar. The projection median of a set of points in $\mathbb{R}^d$. *Discrete and Computational Geometry*, 47(2):329–346, 2011.

[6] P. Bose, A. Maheshwari, and P. Morin. Fast approximations for sums of distances, clustering and the Fermat-Weber problem. *Computational Geometry: Theory and Applications*, 24(3):135–146, 2003.

[7] T. M. Chan. Remarks on k-level algorithms in the plane. Manuscript, 1999.

[8] T. M. Chan. Dynamic planar convex hull operations in near-logarithmic amortized time. *Journal of the ACM*, 48(1):1–12, 2001.

[9] R. Chandrasekaran and A. Tamir. Algebraic optimization: The Fermat-Weber location problem. *Mathematical Programming*, 46:219–224, 1990.

[10] P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, New York, 2^{nd} edition, 1984.

[11] L. Devroye and L. Györfi. *Nonparametric Density Estimation: The L_1 View*. Wiley, 1985.

[12] T. K. Dey. Improved bounds for planar k-sets and related problems. *Discrete and Computational Geometry*, 19:373–382, 1998.

[13] D. L. Donoho and M. Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics*, 20(4):1803–1827, 1992.

[14] S. Durocher. *Geometric Facility Location under Continuous Motion*. PhD thesis, University of British Columbia, 2006.

[15] S. Durocher and D. Kirkpatrick. The projection median of a set of points in $\mathbb{R}^2$. In *Proceedings of the Canadian Conference on Computational Geometry (CCCG)*, volume 17, pages 46–50, 2005.

[16] S. Durocher and D. Kirkpatrick. The Steiner centre: Stability, eccentricity, and applications to mobile facility location. *International Journal of Computational Geometry and Applications*, 16(4):345–371, 2006.

[17] S. Durocher and D. Kirkpatrick. The projection median of a set of points. *Computational Geometry: Theory and Applications*, 42(5):364–375, 2009.

[18] H. Edelsbrunner and E. Welzl. Constructing belts in two-dimensional arrangements with applications. *SIAM Journal on Computing*, 15:271–284, 1986.

[19] S. L. Hakimi. Location theory. In Rosen, Michaels, Gross, Grossman, and Shier, editors, *Handbook of Discrete and Combinatorial Mathematics*. CRC Press, 2000.

[20] P. Indyk. Sublinear time algorithms for metric space problems. In *Proceedings of the Symposium on the Theory of Computing (STOC)*, volume 31, pages 428–434, 1999.

[21] Y. S. Kupitz and H. Martini. Geometric aspects of the generalized Fermat-Torricelli problem. In *Intuitive Geometry*, volume 6, pages 55–127. Bolyai Society Mathematical Studies, Budapest, 1997.

[22] J.-C. Massé. Multivariate trimmed means based on the Tukey depth. *Journal of Statistical Planning and Inference*, 139(2):366–384, 2009.

[23] D. J. Price. Some unusual series occurring in n-dimensional geometry. *The Mathematical Gazette*, 30:149–150, 1946.

[24] Y. V. Prokhorov. An extension of S. N. Bernstein's inequalities to multidimensional distributions. *Theory of Probability and its Applications*, 13(2):260–267, 1968.

[25] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[26] A. Schärlig. About the confusion between the center of gravity and Weber's optimum. *Regional and Urban Economics*, 3(4):371–382, 1973.

[27] G. C. Shephard. Steiner points of convex polytopes. *Canadian Journal of Mathematics*, 18:1294–1300, 1966.

[28] C. G. Small. A survey of multidimensional medians. *International Statistical Review*, 58(3):263–277, 1990.

[29] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Mathematical Journal*, 43:355–386, 1937.

[30] G. O. Wesolowsky. The Weber problem: History and perspectives. *Location Science*, 1(1):5–23, 1993.

[31] Y. Zuo. Multivariate trimmed means based on data depth. In Y. Dodge, editor, *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, Statistics for Industry and Technology, pages 313–322. Birkhauser Basel, 2002.

[32] Y. Zuo and R. Serfling. General notions of statistical depth function. *Annals of Statistics*, 28(2):461–482, 2000.