Efficient Subspace Approximation Algorithms

**Nariankadu D. Shyamalkumar ·
Kasturi Varadarajan**

Abstract We consider the problem of fitting a subspace of a specified dimension k
to a set P of n points in $\mathbb{R}^d$. The fit of a subspace F is measured by the L_τ norm, that
is, it is defined as the τ-root of the sum of the τth powers of the Euclidean distances
of the points in P from F, for some $\tau \geq 1$. Our main result is a randomized algorithm
that takes as input P, k, and a parameter $0 < \varepsilon < 1$; runs in $nd \cdot 2^{O(\frac{\tau k^2}{\varepsilon} \log^2 \frac{k}{\varepsilon})}$ time,
and returns a k-subspace that with probability at least $1/2$ has a fit that is at most
$(1 + \varepsilon)$ times that of the optimal k-subspace.

Keywords Dimension reduction · Low-rank approximation · Shape fitting

1 Introduction

Confronted with high-dimensional data arising from either word-document count,
global climate patterns or any one of the myriad other sources, most scientific ap-
proaches extract a *good* low-dimensional summary. This desire to reduce dimension-
ality may be seen as a consequence of *Occam's Razor*, and the scientific methodolo-
gies we have in mind include those from data mining and statistics.

Computationally efficient statistical techniques for reducing dimensionality in-
clude principal component analysis and multidimensional scaling [24]. These essen-
tially *fit* a flat of a specified dimension to the set of observed points in a relatively

N.D. Shyamalkumar
Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA, USA
e-mail: shyamal-kumar@uiowa.edu

K. Varadarajan (✉)
Department of Computer Science, University of Iowa, Iowa City, IA, USA
e-mail: kvaradar@cs.uiowa.edu

high-dimensional space. In the case that the data points lie *close* to a low-dimensional flat, these techniques are also *consistent* in the sense of being able to discover this flat, asymptotically. Interestingly, these techniques can also be adapted to discover underlying low-dimensional non-linear manifolds, see for example [27, 29, 31]. Motivated by such applications we consider the problem of fitting a flat of a specified dimension to a finite set P of n points in $\mathbb{R}^d$.

Specifically, we are interested in the following *approximate subspace fitting* problem: Given P as above, an integer $0 \le k \le d - 1$, a measure $\mathcal{RD}(F', P)$ of the fit of any subspace F' to P, and a parameter $\varepsilon \ge 0$, find a k-dimensional subspace F such that $\mathcal{RD}(F, P) \le (1 + \varepsilon)\mathcal{RD}(F', P)$ for every k-dimensional subspace F'. We will refer to the special case where $\varepsilon = 0$ as the *exact* subspace fitting problem. Here and throughout the rest of the paper, the word 'subspace' will refer to a linear subspace and the phrase 'k-subspace' will refer to a k-dimensional linear subspace. A *flat* (resp., k-flat) F in $\mathbb{R}^d$ is defined to be a translation of a subspace (resp., k-subspace).

For each $\tau \ge 1$, a measure of how well the flat F fits P is $\mathcal{RD}_\tau(F, P) = (\sum_{p \in P} d(p, F)^\tau)^{1/\tau}$, where $d(p, F)$ is some measure of the distance between point p and flat F. In the case $\tau = \infty$, $\mathcal{RD}_\tau(F, P)$ is naturally defined as $\max_{p \in P} d(p, F)$. In this article, we take $d(p, F)$ to be $\min_{x \in F} \| p - x \|_2$, the minimum Euclidean distance between p and a point in F. We will consider the flat fitting problem with the fit thus measured by $\mathcal{RD}_\tau(\cdot, \cdot)$. This problem has received considerable attention, particularly for the cases $\tau = 1, 2, \infty$. We note that the choice of τ can be very important in some applications. For example, let us consider the case of $k = 0$. The optimal flat with $\tau = 1$ is the *spatial median* which is known to be very robust to the presence of outliers in contrast to the sample mean which corresponds to the case $\tau = 2$, see for example [6] and the references therein. Our main result is that for any $\tau \ge 1$, the approximate flat fitting problem can be solved in $O(nd)$ time, with the constant of proportionality depending solely on ε, k, and τ. Importantly, note that the dimension d is considered part of the input and not a constant.

We now review some work on the flat fitting problem, beginning with the case $\tau = \infty$. When $k = 0$, the problem corresponds to the minimum enclosing ball problem and can be solved in time polynomial in the number of points, the dimension, and $\log \frac{1}{\varepsilon}$; see for instance [17]. The case $k = 1$, the minimum enclosing cylinder problem, is NP-hard [25]. For any fixed k, there are algorithms that solve the problem in $O(nd C_{\varepsilon,k})$ time, where $C_{\varepsilon,k}$ is a constant that is exponential in $2^k/\varepsilon$ [22, 26]. When k is part of the input and is large relative to d, the problem becomes hard to approximate in polynomial time to within a factor of $(\log n)^\delta$, for some $\delta > 0$ [4, 30]. The best known polynomial-time approximation algorithms yield an approximation guarantee of $O(\sqrt{\log n})$ [30]. (For the case $k = d - 1$, an approximation guarantee of $O(\sqrt{\frac{d}{\log d}})$ is also known [5].) If d is a constant, the problem can be solved in time polynomial in the number of points and $\log \frac{1}{\varepsilon}$ for every k [14]; approximation algorithms with running time near linear in n and exponential in d are also known [1].

We now turn to the case $\tau = 2$, focusing on the subspace fitting problem, where some remarkable algebraic properties help the problem. For instance, it is well known that the optimal k-subspace is obtained by the span of the k right singular vectors cor-

responding to the top k singular values of the singular value decomposition (SVD) of the $n \times d$ matrix whose rows correspond to points in P. This leads to a polynomial (in fact, $O(nd \times \min\{n, d\})$) time algorithm for this problem; see the discussion in [11]. For the ε-approximate problem, recent works give algorithms that are near linear in $\frac{ndk}{\varepsilon}$ [2, 10, 19, 28]; this running time improves over $O(nd \times \min\{n, d\})$ when k is small.

The case $\tau = 1$ and $k = 0$ of flat fitting is the *Fermat–Weber* problem, which reduces to minimizing a convex function over $\mathbb{R}^d$. For this problem, an algorithm that is polynomial in the number of points, the dimension, and $\log \frac{1}{\varepsilon}$ is given by [7]. The case $k = d - 1$ is referred to as the median hyperplane problem. Assuming the input point set P spans $\mathbb{R}^d$, it was observed that the optimal hyperplane is the span of a subset of d points of P. Based on this, algorithms that run in $O(n^d)$ time are known for solving this problem exactly; see the surveys [12, 23]. For $0 < k < d - 1$, we are not aware of other work on the polynomial-time solvability of this problem for either the exact or approximate versions. If d is fixed, ε-approximation algorithms that are near linear in n (but exponential in d) are known, see [15, 20].

A problem related to the median hyperplane problem is the well-studied L_1 regression problem; here we are given a real $n \times d$ matrix A, and a column vector $b \in \mathbb{R}^n$, and the goal is to find $x \in \mathbb{R}^d$ that minimizes $\|Ax - b\|_1$. For recent work on this and more generally the L_τ regression we refer to [8, 13].

The flat fitting problem studied in this article assumes the underlying distance $d(p, F)$ of a point p from flat F to be defined in terms of the Euclidean norm. Our techniques appear to rely on the special properties of the Euclidean norm. There has been significant interest in the flat fitting problem when this distance is defined using other norms [17, 18]. We briefly mention a few highlights from [17] here, focusing on the case where $\tau = \infty$. When $d(p, F)$ is defined using the $\| \cdot \|_\infty$ norm, that is, when $d(p, F) = \min_{x \in F} \|p - x\|_\infty$, the flat fitting problem with $k = d - 1$ remains NP-hard. On the other hand, when $d(p, F)$ is defined using the $\| \cdot \|_1$ norm, the flat fitting problem with $k = d - 1$ has a polynomial-time algorithm. Thus the choice of the norm used to measure the distance has significant consequences for the computational complexity. At the other extreme, the flat fitting problem with $k = 0$ has a polynomial-time algorithm when the distance $d(p, F)$ is measured using either the $\| \cdot \|_\infty$ or the $\| \cdot \|_1$ norm. For the case of constant $k \geq 1$, we are not aware of prior work on the subspace fitting problem when the distance $d(p, F)$ is measured using norms other than the Euclidean norm.

1.1 Results and Techniques

Our main result is the following.

Theorem 1.1 *There is a randomized algorithm that, given any set P of n points in $\mathbb{R}^d$, any $1 \leq k < d$, $0 < \varepsilon < 1$, and constant $\tau \in [1, \infty)$, runs in $O(\frac{ndk^2}{\varepsilon} \log \frac{k}{\varepsilon})$ time and returns with probability $2^{-O(\frac{\tau k^2}{\varepsilon} \log^2 \frac{k}{\varepsilon})}$ a k-subspace F such that*

$$\mathcal{RD}_\tau(F, P) \leq (1 + \varepsilon)\mathcal{RD}_\tau(F', P),$$

for any k-subspace F'.

It is also useful to highlight the version of this result where the success probability is at least $1/2$; its proof is included in the Appendix.

Corollary 1.2 *There is a randomized algorithm that, given any set P of n points in $\mathbb{R}^d$, any $1 \le k < d$, $0 < \varepsilon < 1$, and $\tau \in [1, \infty)$, runs in $nd \cdot 2^{O(\frac{\tau k^2}{\varepsilon} \log^2 \frac{k}{\varepsilon})}$ time and returns with probability at least $1/2$ a k-subspace F such that*

$$\mathcal{RD}_\tau(F, P) \le (1 + \varepsilon)\mathcal{RD}_\tau(F', P),$$

for any k-subspace F'.

Thus the running time of the algorithm of Corollary 1.2 is $O(nd)$ if τ, k, and ε are treated as constants, and the constant is exponential in $\frac{\tau k}{\varepsilon}$. Ours is the first work that achieves such a running time for each $\tau \ne 2, \infty$.

We now outline our algorithm for Theorem 1.1 and its analysis, starting with the case $k = 1$, where we want to fit a line. We compute a sequence of lines $\ell_0, \ell_1, \ldots, \ell_i$, where $i = O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$, as follows. We pick a random point from the input set P by sampling according to the τth powers of the Euclidean norms of the points, and let ℓ_0 be the line through this point. Having computed $\ell_0, \ldots, \ell_j$, we extend the sequence by picking a random point w from P according to the same distribution as above, and letting ℓ_{j+1} be a randomly chosen line (according to a suitable distribution) from the 2-subspace spanned by ℓ_j and w. Having computed the entire sequence $\ell_0, \ell_1, \ldots, \ell_i$, we pick one of these $i + 1$ lines uniformly at random and return this as our solution.

We note that our algorithm avoids the curse of dimensionality by generating ℓ_{j+1} from a known 2-subspace. Our analysis begins by observing that ℓ_0 is likely to be close to the optimal line ℓ^* in terms of angle. Furthermore, if $\mathcal{RD}_\tau(\ell_j, P) > (1 + \varepsilon)\mathcal{RD}_\tau(\ell^*, P)$, then it is likely that the sampled point w will be a "witness" to it, that is, $d(w, \ell_j) > (1 + \frac{\varepsilon}{2})d(w, \ell^*)$. If w is in fact a witness, we show that ℓ_{j+1} is likely to make a significantly smaller angle with ℓ^* when compared to ℓ_j.

The argument concludes by noting that if there are sufficiently many angle improvements, then one of the lines in the computed sequence is a solution with the desired accuracy. For $k > 1$, our algorithm works by guessing a larger sequence of $O(\frac{k}{\varepsilon} \log \frac{k}{\varepsilon})$ lines using the same algorithm as above. We show that with probability $2^{-O(\frac{\tau k}{\varepsilon} \log^2 \frac{k}{\varepsilon})}$ at least one line ℓ in the sequence has the property that a k-subspace containing ℓ is nearly optimal. The algorithm then guesses ℓ from this sequence, projects to the orthogonal complement of ℓ, and recursively finds a nearly optimal $(k-1)$-subspace. The algorithm returns the k-subspace spanned by ℓ and this $(k-1)$-subspace.

Theorem 1.1 is readily extended to the approximate flat fitting problem as well. Essentially, the problem of finding an approximately optimal flat of dimension k can be reduced to the problem of finding an approximately optimal subspace of dimension $k + 1$. To keep our exposition from being repetitive, we focus on the subspace fitting problem.

1.2 Discussion of Related Work

Our algorithm and analysis draws ideas from several recent papers. Bădoiu et al. [3] highlighted a useful principle when studying related problems for $k = 0$: if a candidate point is not nearly optimal, then a point in P that is much closer to the optimal point compared to the candidate point can, in some sense, be used to make progress from the current candidate point. Har-Peled and Varadarajan [22], who consider the case $\tau = \infty$, show how this principle can be refined and usefully applied when $k > 0$. This principle in its further refined form plays a role here. Another related idea from Frieze et al. [16] and Bădoiu et al [3] is the possibility of avoiding the curse of dimensionality by working in the span of a small number of appropriately chosen points from P. Finally, Frieze et al. [16] and Deshpande et al. [11], addressing the case of $\tau = 2$, use the idea of sampling points from P in proportion to their squared norms. Our algorithm samples points in proportion to the τth power of their norms to guess the sequence of lines referred to above. Our main contribution is to show that further development of these ideas along with some new ones has the ability to address the approximate flat fitting problem for all $\tau \geq 1$.

Our algorithm and analysis are perhaps closest in spirit to the work of Har-Peled and Varadarajan [22], who address the case of $\tau = \infty$. A key difference is that for the case of $k > 1$, our algorithm does not have a k-subspace at hand that it repeatedly improves upon.

A comparison of the bounds in Corollary 1.2 with the results of [10, 19] for the case $\tau = 2$ is useful. These results rely on Theorem 6 from [11] whose proof exploits the fact that the optimal k-subspace is given by the SVD. Since such a characterization of the optimal k-subspace is lacking for the case $\tau \neq 2$, we have to resort to different methods. Another consequence of the SVD is that it allows the computation of the optimal k-subspace for $\tau = 2$ in $O(nd^2)$ time. If one is able to restrict the search to a space of much smaller dimension, the running time can be improved further. This is the approach that [10, 19, 28] take, enabling a running time that is $nd \operatorname{poly}(\frac{k}{\varepsilon})$.

Building on some of the techniques from this article, Deshpande and Varadarajan [9] have shown that for subspace fitting with any $\tau \geq 1$, one can compute in $nd \operatorname{poly}(\frac{\tau k}{\varepsilon})$ time a subspace V of dimension $\operatorname{poly}(\frac{\tau k}{\varepsilon})$ that with high probability contains an approximately optimal k-subspace. This is a dimension reduction result that allows us to restrict our search for a nearly optimal k-subspace to within the low-dimensional subspace V. Such a search can be conducted by plugging in modifications of algorithms developed for the fixed-dimensional context, such as the methods of Har-Peled [20], because the dimension of the ambient space is now effectively $\operatorname{poly}(\frac{\tau k}{\varepsilon})$. However, this approach does not lead to significant improvements over the running time guaranteed by Theorem 1.1. In particular the exponential dependence on k/ε remains, because all previous methods such as that of [20] have at least an exponential dependence on the dimension when applied to the subspace fitting problem for $\tau \neq 2$.

Organization of the Paper   The remainder of this paper is organized as follows. In Sect. 2, we describe some useful geometric lemmas and a definition. In Sect. 3, we

show that for any set P of points in $\mathbb{R}^d$ and $\varepsilon > 0$, there exists a subset $Q \subseteq P$ with size $O(\frac{k^2}{\varepsilon} \log \frac{k}{\varepsilon})$ whose span contains an approximately optimal k-subspace for P. This Q can be thought of as a "core-set" for the problem, in the sense that the search for an approximately optimal subspace can be restricted to the span of Q, whose dimension is bounded by the size of Q and therefore independent of n and d. Notice that this is a rather weak requirement of a core-set. A stronger notion, for instance, is to require that a nearly optimal k-subspace for Q is also nearly optimal for the original set P. This paper does not construct core-sets satisfying these stronger conditions. Core-sets (of different flavors) have played an important role in the development of fast approximation algorithms for several geometric problems [1].

Although our core-set result is implicit (and thus subsumed) in the proof of Theorem 1.1, we present a direct and much simpler proof that serves as motivation and warm-up for the algorithm of Theorem 1.1. Core-sets with sizes bounds similar to ours were shown for the approximate subspace fitting problem with $\tau = \infty$ and $\tau = 2$ by Har-Peled and Varadarajan [21] and Deshpande et al. [11], respectively. Such core-sets have applications to the *projective clustering* problem, where we want to fit multiple subspaces. For more details we refer to [11, 21]. The core-set result is only existential, in the sense that it does not directly yield an algorithm for computing the core-set within the time bounds of Theorem 1.1. In Sect. 4, we develop the algorithm that establishes Theorem 1.1, our main result.

2 Preliminaries

The following lemma is taken from [22] and is given below for the convenience of the reader. For brevity, we will use $\|v\|$ to denote the 2-norm $\|v\|_2$ of a point $v \in \mathbb{R}^d$.

Lemma 2.1 *Let w, w^*, and w' be three points in $\mathbb{R}^d$ such that $\|w - w'\| \geq (1 + \varepsilon)\|w - w^*\|$, where $0 < \varepsilon \leq 1$. Then there exists a point s on the segment $\overline{w'w}$ such that*

$$\|w^* - s\| \leq (1 - \varepsilon/2)\|w^* - w'\|.$$

Proof Let $\rho = \|w - w^*\|/\|w - w'\|$, and let s be the point on the segment $\overline{w'w}$ at a distance $\rho\|w - w^*\|$ from w (see Fig. 1). It is easy to see that the triangle wsw^* is similar to the triangle ww^*w' with a scaling factor of ρ. Therefore,
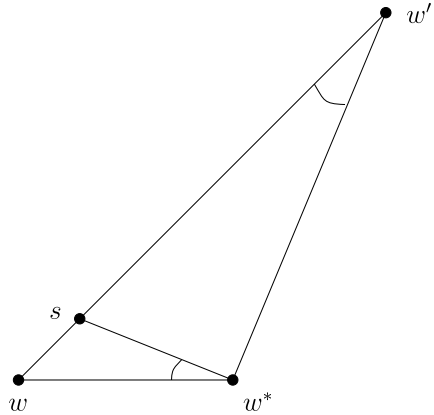
$$\|s - w^*\| = \rho\|w^* - w'\| \leq \|w^* - w'\|/(1 + \varepsilon) \leq (1 - \varepsilon/2)\|w^* - w'\|. \qquad \square$$

The following lemma is in the same spirit as Lemma 2.4 of [22]. Recall that $d(p, F) = \min_{x \in F} \|p - x\|$ denotes the distance of point p from subspace F.

Lemma 2.2 *Let $r, v \in \mathbb{R}^d$ be any two points, let f, z be any two points on the line through r and v, and G a subspace. Then*

$$\frac{|d(z, G) - d(f, G)|}{\|z - f\|} \leq \frac{d(r, G) + d(v, G)}{\|r - v\|}.$$

Fig. 1 The triangles ww^*w' and wsw^* are similar, because $\frac{\|w-w^*\|}{\|w-w'\|} = \frac{\|w-s\|}{\|w-w^*\|}$

Proof For a point $a \in \mathbb{R}^d$, let a_G and $a_{G^\perp}(= a - a_G)$ be the projections of a onto G and $G^\perp$, respectively, where $G^\perp$ is the orthogonal complement of G. Also note that $d(a, G) = \|a_{G^\perp}\|$. We have

$$\frac{|d(z, G) - d(f, G)|}{\|z - f\|} = \frac{|\|z_{G^\perp}\| - \|f_{G^\perp}\||}{\|z - f\|} \leq \frac{\|z_{G^\perp} - f_{G^\perp}\|}{\|z - f\|}$$

$$= \frac{\|r_{G^\perp} - v_{G^\perp}\|}{\|r - v\|} \leq \frac{\|r_{G^\perp}\| + \|v_{G^\perp}\|}{\|r - v\|}$$

$$= \frac{d(r, G) + d(v, G)}{\|r - v\|}. \qquad \square$$

Let us recall that for each $\tau \geq 1$, our measure of how well the subspace F fits P is $\mathcal{RD}_\tau(F, P) = (\sum_{p \in P} d(p, F)^\tau)^{1/\tau}$. We conclude this section by defining the rotation of a k-subspace F through a line ℓ that passes through the origin.

Definition 1 Let F be a k-subspace and ℓ a line through the origin. If the projection of a line ℓ onto F is the origin, then we take any $(k-1)$-subspace H of F, and define the rotation of F through ℓ to be the k-subspace spanned by H and ℓ. Otherwise, the projection of ℓ onto F is a line ℓ'. In this case we take H to be the orthogonal complement of ℓ' in F, and define the rotation of F through ℓ to be the k-subspace spanned by H and ℓ.

3 Small Core-Sets for *k*-Subspaces

In this section, we show that any point set P has a small subset whose span contains an approximately optimal k-subspace:

Theorem 3.1 *Fix $\tau \geq 1$. Let P be a set of points in $\mathbb{R}^d$ and let F^* be the k-subspace that minimizes $\mathcal{RD}_\tau(\cdot, P)$ over all k-subspaces. In the nontrivial case of strictly positive $\mathcal{RD}_\tau(F^*, P)$, and any $0 < \varepsilon < 1$, there exists a $Q \subset P$ consisting of $O(\frac{k^2}{\varepsilon} \log \frac{k}{\varepsilon})$ points such that the span of Q contains a k-subspace F such that*

$$\mathcal{RD}_\tau(F, P) \leq (1 + \varepsilon)\mathcal{RD}_\tau(F^*, P).$$

The proof of Theorem 3.1 is composed of two parts: in Lemma 3.2, we first establish it for the case $k = 1$ and in Lemma 3.3, we leverage this to the case of larger k via an inductive argument. The overall approach in the proof resembles that of [21] for the case $\tau = \infty$; certain simplifications and improvements (that pertain to Lemma 3.2) enable us to handle any $\tau \geq 1$.

3.1 Core-Sets for Lines

Lemma 3.2 *Let P be a set of points in $\mathbb{R}^d$, and let ℓ^* be the line that minimizes $\mathcal{RD}_\tau(\ell, P)$ over all lines ℓ through the origin o. Assume that $\mathcal{RD}_\tau(\ell^*, P) > 0$. There exists a constant $c > 0$ such that for any $0 < \varepsilon < 1$, there exists a set of at most $\frac{c}{\varepsilon} \log \frac{1}{\varepsilon}$ points in P whose span contains a point $t \neq o$ such that $\mathcal{RD}_\tau(\ell(t), P) \leq (1 + \varepsilon)\mathcal{RD}_\tau(\ell^*, P)$, where $\ell(p)$ denotes the line through p (assumed different from o) and o.*

Proof For any $p \in \mathbb{R}^d$, let p^* denote its projection onto ℓ^*. For any line ℓ through the origin, let $\alpha(\ell)$ denote the distance $\|x - x^*\|$, where x is a point on ℓ at distance 1 from o. Note that $\alpha(\ell)$ stands for the sine of the angle between ℓ and ℓ^*.

We construct a sequence $q_0, \ldots, q_i$ of points different from o, where $i = \lceil \frac{c}{\varepsilon} \log \frac{1}{\varepsilon} \rceil$ and $c > 0$ is an appropriate constant, such that

1. q_0 is the point in P that minimizes $\alpha(\ell(p))$ over each $p \in P$ distinct from o.
2. q_j is in the span of at most $j + 1$ points from P.
3. For $1 \leq j \leq i$, if

$$\mathcal{RD}_\tau(\ell(q_{j-1}), P) > (1 + \varepsilon)\mathcal{RD}_\tau(\ell^*, P),$$

   then $\alpha(\ell(q_j)) \leq (1 - \varepsilon/2)\alpha(\ell(q_{j-1}))$.

The sequence consisting of just the point q_0 clearly satisfies conditions (2) and (3). Suppose that we have inductively constructed the sequence $q_0, \ldots, q_{j-1}$, where $j \geq 1$. We describe how to extend the sequence. If $\mathcal{RD}_\tau(\ell(q_{j-1}), P) \leq (1 + \varepsilon)\mathcal{RD}_\tau(\ell^*, P)$, this is trivial because we can take q_j to be any point of P different from o. Otherwise, there is a point $w \in P$ such that $d(w, \ell(q_{j-1})) > (1 + \varepsilon)d(w, \ell^*)$. Let w' denote the projection of w^* onto $\ell(q_{j-1})$. We have $\|w - w'\| > (1 + \varepsilon)\|w - w^*\|$. From Lemma 2.1, there is a point s on the segment $\overline{w'w}$ (see Fig. 2) such that $\|w^* - s\| \leq (1 - \varepsilon/2)\|w^* - w'\|$. Let $q_j = s$. Since $\|w^* - q_j\| < \|w^* - w'\| \leq \|w^* - o\|$, q_j is different from o. We have

$$d(w^*, \ell(q_j)) \leq \|w^* - s\| \leq (1 - \varepsilon/2)\|w^* - w'\| = (1 - \varepsilon/2)d(w^*, \ell(q_{j-1})).$$
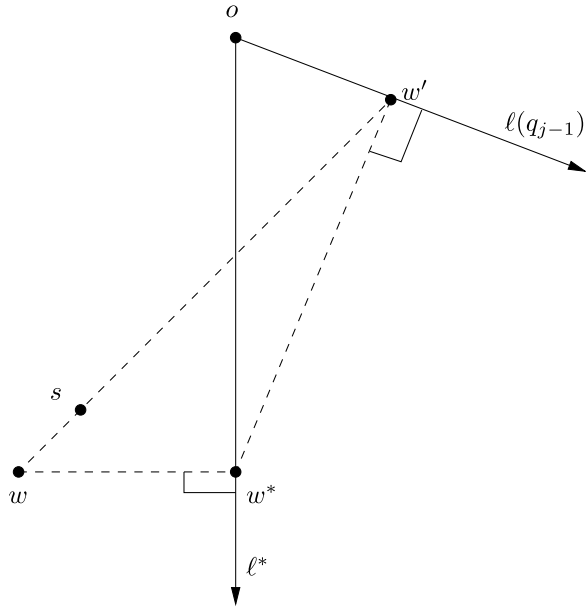
This implies that $\alpha(\ell(q_j)) \leq (1 - \varepsilon/2)\alpha(\ell(q_{j-1}))$, and so the extended sequence satisfies condition (3). Since q_{j-1} is in the span of at most j points from P, and q_j lies in the span of q_{j-1} and w, q_j lies in the span of at most $j + 1$ points from P. So the extended sequence also satisfies condition (2).

We now argue that at least one of the points q_j in the sequence satisfies $\mathcal{RD}_\tau(\ell(q_j), P) \leq (1+\varepsilon)\mathcal{RD}_\tau(\ell^*, P)$, thus proving the lemma. If $\ell(q_0), \ldots, \ell(q_{i-1})$ do not satisfy this inequality then condition (3) tells us that

$$\alpha\big(\ell(q_i)\big) \leq (1 - \varepsilon/2)^i \alpha\big(\ell(q_0)\big) \leq e^{-\varepsilon i/2} \alpha\big(\ell(q_0)\big) \leq e^{-\log\frac{1}{\varepsilon}} \alpha\big(\ell(q_0)\big) = \varepsilon\alpha\big(\ell(q_0)\big)$$

by our choice of $i = \lceil \frac{c}{\varepsilon} \log \frac{1}{\varepsilon} \rceil$, provided $c \geq 2$. Then for any $p \in P$, we have

$$d\big(p, \ell(q_i)\big) \leq \|p - p^*\| + \|p^*\|\alpha\big(\ell(q_i)\big) \leq \|p - p^*\| + \|p\|\varepsilon\alpha\big(\ell(q_0)\big)$$

$$\leq \|p - p^*\| + \varepsilon\|p\|\frac{\|p - p^*\|}{\|p\|} = (1 + \varepsilon)\|p - p^*\| = (1 + \varepsilon)d\big(p, \ell^*\big),$$

where the third inequality follows from the choice of q_0. It then follows that $\mathcal{RD}_\tau(\ell(q_i), P) \leq (1 + \varepsilon)\mathcal{RD}_\tau(\ell^*, P)$.   □

3.2 Core-Sets for k-Subspaces

Lemma 3.3 *Let P be a set of points in $\mathbb{R}^d$ and let F^* be the k-subspace that minimizes $\mathcal{RD}_\tau(F, P)$ over all k-subspaces F, where $k \geq 1$. Assume that $\mathcal{RD}_\tau(F^*, P) > 0$. Let $0 < \varepsilon < 1$ be a parameter, and let $\delta = \varepsilon/2k$. There exists a set of at most $\frac{ck}{\delta} \log \frac{1}{\delta} = \frac{ck^2}{\varepsilon} \log \frac{k}{\varepsilon}$ points in P whose span contains a k-subspace G*

such that $\mathcal{RD}_\tau(G, P) \leq (1+\delta)^k \mathcal{RD}_\tau(F^*, P) < (1+\varepsilon)\mathcal{RD}_\tau(F^*, P)$. Here $c > 0$ is the constant appearing in Lemma 3.2.

Proof The proof, which resembles that of Lemma 2.4 of [21], is by induction on k. The base case of $k = 1$ is furnished by Lemma 3.2. So we suppose that $k > 2$.

Let $e_1, \ldots, e_k$ denote a set of orthogonal unit vectors on F^*. Let π denote the projection to the orthogonal complement of the subspace spanned by $e_1, \ldots, e_{k-1}$. The crucial property of π is that for any k-subspace F in $\mathbb{R}^d$ that contains $e_1, \ldots, e_{k-1}$ and any point p, we have $d(p, F) = d(\pi(p), \pi(F))$. Note that for such a k-subspace F, $\pi(F)$ is a line.

Using Lemma 3.2, there exists a set $Q_1 \subseteq P$ of at most $\frac{c}{\delta} \log \frac{1}{\delta}$ points such that the span of $\pi(Q_1)$ contains a line ℓ such that

$$\mathcal{RD}_\tau\big(\ell, \pi(P)\big) \leq (1+\delta)\mathcal{RD}_\tau\big(\pi\big(F^*\big), \pi(P)\big).$$

Let F be the k-subspace in $\mathbb{R}^d$ spanned by ℓ and $e_1, \ldots, e_{k-1}$. The key property of π implies that

$$\mathcal{RD}_\tau(F, P) \leq (1+\delta)\mathcal{RD}_\tau\big(F^*, P\big).$$

Using the linearity of π and its key property, we can also conclude that there is a line ℓ' through o that is contained in F as well as in the span of Q_1.

Now consider the projection π' to the orthogonal complement of ℓ'. We have

$$\mathcal{RD}_\tau\big(\pi'(F), \pi'(P)\big) = \mathcal{RD}_\tau(F, P),$$

$\pi'(F)$ is a $(k-1)$-subspace, and $\mathcal{RD}_\tau(H, \pi'(P)) > 0$ for any $(k-1)$-subspace H in the orthogonal complement of ℓ'. Inductively, we obtain a set $Q_2 \subseteq P$ of at most $\frac{(k-1)c}{\delta} \log \frac{1}{\delta}$ points such that the span of $\pi'(Q_2)$ contains a $(k-1)$-subspace H such that $\mathcal{RD}_\tau(H, \pi'(P)) \leq (1+\delta)^{k-1}\mathcal{RD}_\tau(\pi'(F), \pi'(P))$. Let G be the k-subspace such that $\pi'(G) = H$. That is, G is the span of H and ℓ'. We have $\mathcal{RD}_\tau(G, P) = \mathcal{RD}_\tau(H, \pi'(P))$ and so

$$\mathcal{RD}_\tau(G, P) \leq (1+\delta)^{k-1}\mathcal{RD}_\tau\big(\pi'(F), \pi'(P)\big)$$

$$= (1+\delta)^{k-1}\mathcal{RD}_\tau(F, P)$$

$$\leq (1+\delta)^k \mathcal{RD}_\tau\big(F^*, P\big).$$

Since ℓ' lies in the span of Q_1, and H lies in the span of $\pi'(Q_2)$, we can conclude that G lies in the span of $Q = Q_1 \cup Q_2$. Clearly, $|Q| \leq |Q_1| + |Q_2| \leq \frac{ck}{\delta} \log \frac{1}{\delta}$. $\square$

4 Efficient Computation of Good Subspaces

In this section, we describe the algorithm and the analysis needed to establish Theorem 1.1.

4.1 The Algorithm

Let $\delta = \varepsilon/2k$; note that $(1 + \delta)^k \leq (1 + \varepsilon)$. We now describe a recursive algorithm, Good-Subspace, that takes as arguments a subspace $\mathcal{S}$ of $\mathbb{R}^d$, a (multi-) set P of points lying on $\mathcal{S}$, an integer $1 \leq k < \dim(\mathcal{S})$, and the parameter δ. Let F^* denote the k-subspace in $\mathcal{S}$ that minimizes $\mathcal{RD}_\tau(\cdot, P)$. The algorithm Good-Subspace returns a k-subspace $\hat{F}_k$ of $\mathcal{S}$, and we will later argue that $\mathcal{RD}_\tau(\hat{F}_k, P) \leq (1+\delta)^k \mathcal{RD}_\tau(F^*, P)$ with a reasonably large probability. The parameter on which the algorithm recurses is k; the base case will be $k = 1$. (The algorithm for Theorem 1.1 is obtained by calling Good-Subspace on inputs $\mathcal{S} = \mathbb{R}^d$, P, k, and δ.)

If every point in the multiset P is the same as o, the origin, we return any k-subspace lying in $\mathcal{S}$. Otherwise, we first compute a sequence $\ell_0, \ldots, \ell_i$ of lines, where $i = \lceil \frac{c}{\delta} \log \frac{1}{\delta} \rceil$ and $c > 0$ is an appropriately chosen constant. The sequence is not deterministic, but a function of the probabilistic choices made by the algorithm. We first pick a random point p from P so that the probability of picking $q \in P$ is $\frac{\|q\|^\tau}{\sum_{p \in P} \|p\|^\tau}$ and set $\ell_0 = \ell(p)$. (Recall that $\ell(p)$ is the line through the origin and p.) Having picked $\ell_0, \ldots, \ell_j$, where $0 \leq j \leq i - 1$, we pick ℓ_{j+1} as follows. We pick a random point w from P according to the same distribution used above. Let u and v be unit vectors in the direction ℓ_j and $\ell(w)$, respectively. We choose one of the following two segments with equal probability: the segment $\overline{uv}$ and the segment $\overline{(-u)v}$. We then pick a point uniformly at random from the chosen segment, and let ℓ_{j+1} be the line through o and the chosen point.

Having computed the sequence $\ell_0, \ldots, \ell_i$, we pick a line ℓ uniformly at random from this sequence.

If $k = 1$, we simply return the line ℓ. Otherwise, let $\mathcal{S}'$ denote the orthogonal complement of ℓ in $\mathcal{S}$. Let π denote the projection function onto $\mathcal{S}'$. We recursively call Good-Subspace with the parameters $\mathcal{S}'$, $\pi(P)$, $k-1$, and δ. The recursive call returns a $(k-1)$-subspace G of $\mathcal{S}'$. The subspace G and ℓ together span a k-subspace of $\mathcal{S}$. This is what the algorithm Good-Subspace returns on inputs $\mathcal{S}$, P, k, and δ.

4.2 Running Time

It is clear that the computation of each line in the sequence can be done in $O(nd)$ time. It also takes $O(nd)$ time to set up the recursive call once we have ℓ. Thus the running time, excluding the time taken by the recursive call, is $O(\frac{nd}{\delta} \log \frac{1}{\delta})$. Since the depth of the recursion is k, the overall running time of the algorithm is $O(\frac{ndk}{\delta} \log \frac{1}{\delta})$.

4.3 Performance

Let F_j denote the rotation of F^* through ℓ_j, for $1 \leq j \leq i$, and $\hat{F}_1$ the rotation of F^* through ℓ. Let $\mu = (\delta/4)^\tau \cdot \delta^2/432$. The following lemma is the essence of the performance guarantee of the algorithm.

Lemma 4.1 *Suppose that P contains some point that is different from o. With a probability of at least $\mu^i/2$, there exists a j between 0 and i such that $\mathcal{RD}_\tau(F_j, P) \leq$*

$(1 + \delta)\mathcal{RD}_\tau(F^*, P)$. *Consequently, we have*

$$\Pr\left(\mathcal{RD}_\tau\left(\hat{F}_1, P\right) \leq (1 + \delta)\mathcal{RD}_\tau\left(F^*, P\right)\right) \geq \frac{\mu^i}{2(i + 1)}.$$

Proof For any $p \in \mathbb{R}^d$, let p^* denote its projection onto F^*. For any line ℓ through the origin, let $\alpha(\ell)$ denote the sine of the angle between ℓ and F^*. That is, $\alpha(\ell) = \|p - p^*\|/\|p\|$ for any point $p \neq o$ on ℓ. For better readability we structure the proof around two claims:

Claim 1 With a probability of at least $1/2$, we have

$$\left(\sum_{p \in P} \|p\|^\tau\right)^{1/\tau} \alpha(\ell_0) \leq 2^{1/\tau} \mathcal{RD}_\tau\left(F^*, P\right). \tag{1}$$

Claim 2 For any $1 \leq j \leq i$, suppose $\mathcal{RD}_\tau(F_{j-1}, P) > (1+\delta)\mathcal{RD}_\tau(F^*, P)$ and suppose $\ell_0, \dots, \ell_{j-1}$ are such that

- ℓ_0 satisfies the inequality (1), and
- $\alpha(\ell_{j'}) \leq \alpha(\ell_{j'-1})$, for $1 \leq j' \leq j - 1$.

Then the probability that $\alpha(\ell_j) \leq (1 - \delta/20)\alpha(\ell_{j-1})$ is at least μ.

Assuming the two claims, it follows that with a probability of at least $\mu^i/2$, the following two events *simultaneously* occur:

1.
$$\left(\sum_{p \in P} \|p\|^\tau\right)^{1/\tau} \alpha(\ell_0) \leq 2\mathcal{RD}_\tau\left(F^*, P\right).$$

2. For each $1 \leq j \leq i$, either

$$\mathcal{RD}_\tau(F_{j-1}, P) \leq (1 + \delta)\mathcal{RD}_\tau\left(F^*, P\right)$$

or

$$\alpha(\ell_j) \leq (1 - \delta/20)\alpha(\ell_{j-1}).$$

We argue that these events imply that

$$\mathcal{RD}_\tau(F_j, P) \leq (1 + \delta)\mathcal{RD}_\tau\left(F^*, P\right)$$

for some j between 0 and i. If this inequality holds for some $0 \leq j \leq i - 1$, we are done. Otherwise, we have

$$\alpha(\ell_i) \leq (1 - \delta/20)^i \alpha(\ell_0) \leq \frac{\delta}{2}\alpha(\ell_0)$$

by our choice of i. Denoting by p' the projection of p^* onto F_i, using Minkowski's inequality and the occurrence of the first event, we have

$$
\mathcal{RD}_\tau(F_i, P) \leq \left(\sum_{p \in P} \left(\|p - p^*\| + \|p^* - p'\| \right)^\tau \right)^{1/\tau}
$$

$$
\leq \left(\sum_{p \in P} \|p - p^*\|^\tau \right)^{1/\tau} + \left(\sum_{p \in P} \|p^* - p'\|^\tau \right)^{1/\tau}
$$

$$
\leq \mathcal{RD}_\tau\left(F^*, P\right) + \alpha(\ell_i) \left(\sum_{p \in P} \|p^*\|^\tau \right)^{1/\tau}
$$

$$
\leq \mathcal{RD}_\tau\left(F^*, P\right) + \frac{\delta}{2} \alpha(\ell_0) \left(\sum_{p \in P} \|p\|^\tau \right)^{1/\tau}
$$

$$
\leq (1 + \delta) \mathcal{RD}_\tau\left(F^*, P\right).
$$

Now all that is left is to prove the claims.

Proof of Claim 1 Let $\mathbb{E}(X)$ denote, as usual, the expectation of random variable X. We have

$$
\left(\sum_{p \in P} \|p\|^\tau \right) \mathbb{E}\left(\alpha(\ell_0)^\tau\right) = \left(\sum_{p \in P} \|p\|^\tau \right) \sum_{q \in P} \frac{\|q\|^\tau}{\sum_{p \in P} \|p\|^\tau} \frac{\|q - q^*\|^\tau}{\|q\|^\tau}
$$

$$
= \sum_{q \in P} \|q - q^*\|^\tau = \mathcal{RD}_\tau\left(F^*, P\right)^\tau.
$$

Non-negativity of $\alpha(\ell)$ implies that with Markov's Inequality we have Claim 1.   □

Proof of Claim 2 Let us call a point $p \in P$ a *witness* if $d(p, F_{j-1}) > (1 + \delta/2) \times d(p, F^*)$ and let P_{j-1} be the set of all witnesses. The proof follows along the following steps with Fig. 3 serving as a visual aid for the argument.

Step 1: We show that there is a sufficient probability of the sampled point w, used to go from ℓ_{j-1} to ℓ_j, being a *witness*.
Step 2: In the case that w is a witness, we show that there is a line $\ell(q)$ through a point q on the span of ℓ_{j-1} and the sampled point w that makes a significantly smaller angle with F^* than ℓ_{j-1}.
Step 3: Our final step is to show that the ℓ_j chosen by the algorithm is close enough to $\ell(q)$ with a reasonable probability.

We proceed to establish the goal of Step 1. Precisely, we prove by contradiction that

$$
\left(\sum_{p \in P_{j-1}} \|p\|^\tau \right)^{1/\tau} \geq \frac{\delta}{4} \left(\sum_{p \in P} \|p\|^\tau \right)^{1/\tau}. \tag{2}
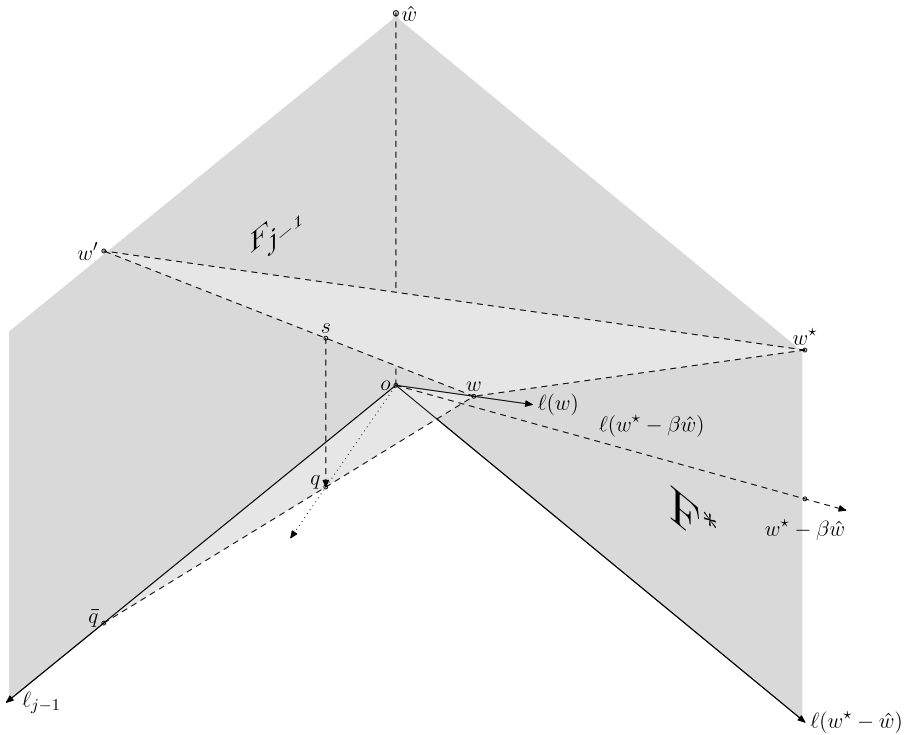$$

Fig. 3 Visual aid for Step 2: The angle $\alpha(\ell(q))$ is significantly smaller than $\alpha(\ell_{j-1})$

which implies that the point w chosen by the algorithm in constructing ℓ_j from ℓ_{j-1} has a probability of at least $(\delta/4)^\tau$ of being a *witness*.

Assuming the contrary of (2), bounding $d(p, F_{j-1})$ by $(1 + \delta/2)d(p, F^*)$ for $p \in P \setminus P_{j-1}$ and by $\|p - p^*\| + d(p^*, F_{j-1}) \leq (1 + \delta/2)d(p, F^*) + d(p^*, F_{j-1})$ for $p \in P_{j-1}$, we have, by Minkowski's inequality,

$$
\left(\sum_{p \in P} d(p, F_{j-1})^\tau \right)^{1/\tau}
$$

$$
\leq (1 + \delta/2) \left(\sum_{p \in P} d(p, F^*)^\tau \right)^{1/\tau} + \left(\sum_{p \in P_{j-1}} d(p^*, F_{j-1})^\tau \right)^{1/\tau}
$$

$$
\leq (1 + \delta/2) \mathcal{RD}_\tau(F^*, P) + \alpha(\ell_{j-1}) \left(\sum_{p \in P_{j-1}} \|p^*\|^\tau \right)^{1/\tau}
$$

$$
\leq (1 + \delta/2) \mathcal{RD}_\tau(F^*, P) + \alpha(\ell_0) \left(\sum_{p \in P_{j-1}} \|p\|^\tau \right)^{1/\tau}
$$

$$\leq (1 + \delta/2)\mathcal{RD}_\tau\big(F^*, P\big) + \frac{\delta}{4}\alpha(\ell_0)\left(\sum_{p \in P} \|p\|^\tau\right)^{1/\tau}$$

$$\leq (1 + \delta/2)\mathcal{RD}_\tau\big(F^*, P\big) + (\delta/2)\mathcal{RD}_\tau\big(F^*, P\big)$$

$$= (1 + \delta)\mathcal{RD}_\tau\big(F^*, P\big).$$

Here, the second inequality is based on the fact that F_{j-1} is the rotation of F^* through ℓ_{j-1} and $\alpha(\ell_{j-1})$ is the sine of the angle between the two subspaces; the penultimate inequality used the fact that ℓ_0 satisfies inequality (1). But we have now arrived at a contradiction to the assumption that $\mathcal{RD}_\tau(F_{j-1}, P) > (1 + \delta)\mathcal{RD}_\tau(F^*, P)$, hence proving (2).

Assuming that w is a *witness*, we proceed to Step 2. Recall that F_{j-1} is the rotation of F^* through ℓ_{j-1}. Let H denote the $(k-1)$-subspace of F_{j-1} and F^* that is used in the definition of the rotation. Observe that H is the orthogonal complement of ℓ_{j-1} in F_{j-1} and also in F^*, the latter holding provided the projection of ℓ_{j-1} onto F^* is a line. Let $\pi_H(\cdot)$ denote the projection onto H. Of course, $\pi_H(\ell_{j-1})$ is just the origin o.

Let w' denote the projection of w^* onto F_{j-1}, where w^*, recall, is the projection of w onto F^*. Since w is a witness, we have $\|w - w'\| > (1 + \delta/2)\|w - w^*\|$. From Lemma 2.1, there is a point s on the segment $\overline{w'w}$ such that $\|w^* - s\| \leq (1 - \delta/4)\|w^* - w'\|$. Let β be such that $s = \beta w' + (1 - \beta)w$.

Let $\hat{w} = \pi_H(w) = \pi_H(w^*) = \pi_H(w')$. We note that the point $\bar{q} = w' - \hat{w}$ lies on the line ℓ_{j-1}, and define q as the point on the segment $\overline{\bar{q}w}$ given by $s - \beta\hat{w} = \beta(w' - \hat{w}) + (1 - \beta)w = \beta\bar{q} + (1 - \beta)w$. We now have

$$\alpha\big(\ell(q)\big) \leq \frac{\|w^* - \beta\hat{w} - q\|}{\|w^* - \beta\hat{w}\|} = \frac{\|w^* - s\|}{\|w^* - \beta\hat{w}\|} \leq \left(1 - \frac{\delta}{4}\right)\frac{\|w^* - w'\|}{\|w^* - \beta\hat{w}\|}$$

$$\leq \left(1 - \frac{\delta}{4}\right)\frac{\|r' - \beta\hat{r} - (\bar{r} - \beta\hat{r})\|}{\|\bar{r} - \beta\hat{r}\|} \leq \left(1 - \frac{\delta}{4}\right)\frac{\|w^* - w'\|}{\|w^* - \hat{w}\|}$$

$$= \left(1 - \frac{\delta}{4}\right)\alpha(\ell_{j-1}). \tag{3}$$

The first inequality follows because we can bound the sine of the angle between $\ell(q)$ and F^* by the sine of the angle between $\ell(q)$ and the line through $w^* - \beta\hat{w}$, which lies on F^*. The penultimate inequality is because $\|w^* - \beta\hat{w}\| = \|(w^* - \hat{w}) + (1 - \beta)\hat{w}\| \geq \|w^* - \hat{w}\|$, as $(w^* - \hat{w}) \cdot \hat{w} = 0$. This guarantee on $\alpha(\ell(q))$ brings us to the end of Step 2.

Now we proceed to our final step and show that the ℓ_j chosen by the algorithm is close to $\ell(q)$ with a reasonable probability. Following the notation of the algorithm, let u and $-u$ denote the unit vectors lying on ℓ_{j-1}, and v the unit vector $\frac{w}{\|w\|}$. Suppose that the inner product $w \cdot u > 0$, and that in fact the angle uow is at most $\pi/4$. In this case, we argue that $\bar{q}$ lies on the ray $\{tu | t > 0\}$. First, observe that $w^* \cdot u > 0$. For otherwise, $\|w - w^*\|$ is at least as large as the distance of w from the hyperplane normal to ℓ_{j-1} which in turn is at least as large as $d(w, \ell_{j-1})$ due to the angle uow being at
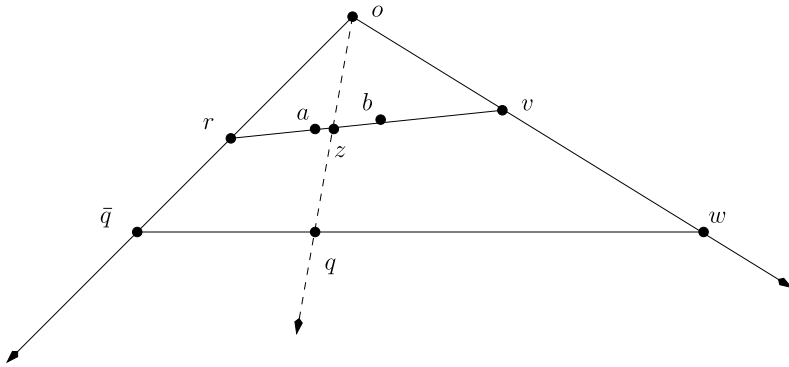
Fig. 4 Visual aid for Step 3. For any f on segment $\overline{ab}$, $\alpha(\ell(f))$ is close enough to $\alpha(\ell(z))$

most $\pi/4$. Hence we have $d(w, F^*) = \|w - w^*\| \geq d(w, \ell_{j-1}) \geq d(w, F_{j-1})$, contradicting the fact that w is a witness. Now since the vectors $w' - w^*$ and $\bar{q} - w'$ are orthogonal to ℓ_{j-1}, we have

$$\bar{q} \cdot u = \big(\bar{q} + (w' - \bar{q}) + (w^* - w')\big) \cdot u = w^* \cdot u > 0,$$

that is, $\bar{q}$ lies on the ray $\{tu | t > 0\}$.

Let $r = u/-u$ be the unit vector such that $\bar{q}$ lies on the ray $\{tr | t \geq 0\}$. Please see Fig. 4 for a visual aid to the argument that follows. We have just argued that the angle rov is at most $3\pi/4$. Since q lies on the segment $\overline{\bar{q}w}$, the line $\ell(q)$ intersects the segment $\overline{rv}$ at some point, call it z. There is a segment $\overline{ab}$ containing z and contained in $\overline{rv}$ so that

$$\frac{\|a - b\|}{\|r - v\|} = \frac{\delta^2}{216}.$$

Since w is a witness, we have $\|w - w^*\| + \|w^* - w'\| \geq \|w - w'\| > (1 + \frac{\delta}{2})\|w - w^*\|$, which implies that $\frac{\delta}{2}\|w - w^*\| \leq \|w^* - w'\|$. Thus,

$$\frac{\delta}{2}\|w\|\alpha\big(\ell(w)\big) = \frac{\delta}{2}\|w - w^*\| \leq \|w^* - w'\| = \|w^* - \hat{w}\| \frac{\|w^* - w'\|}{\|w^* - \hat{w}\|}$$

$$= \|w^* - \hat{w}\|\alpha(\ell_{j-1}) \leq \|w^*\|\alpha(\ell_{j-1}) \leq \|w\|\alpha(\ell_{j-1}),$$

so $\alpha(\ell(w)) \leq \frac{2}{\delta}\alpha(\ell_{j-1})$. Using Lemma 2.2, we see that if f is a point on the segment $\overline{ab}$, then

$$\|f - f^*\| - \|z - z^*\|$$

$$\leq \frac{\|f - z\|}{\|v - r\|}\big(\|v - v^*\| + \|r - r^*\|\big) \leq \frac{\|a - b\|}{\|v - r\|}\big(\|v - v^*\| + \|r - r^*\|\big)$$

$$= \frac{\|a - b\|}{\|v - r\|}\big(\alpha\big(\ell(w)\big) + \alpha(\ell_{j-1})\big) \leq \frac{\delta^2}{216}\left(\frac{4}{\delta}\right)\alpha(\ell_{j-1}) = \frac{\delta}{54}\alpha(\ell_{j-1}).$$

Observe that $\|a - b\| \leq \frac{\delta^2}{216}\|v - r\| \leq \frac{\delta^2}{108}$. Also, for any point f on $\overline{rv}$, we have $\|f\| \geq 1/3$ as a consequence of the angle vor being at most $3\pi/4$. Thus for any $f \in \overline{ab}$, we have

$$\|f\| \geq \|z\| - \|f - z\| \geq \|z\| - \frac{\delta^2}{108} \geq \|z\|\left(1 - \frac{\delta^2}{36}\right).$$

So for any $f \in \overline{ab}$,

$$
\begin{aligned}
\alpha\big(\ell(f)\big) - \alpha\big(\ell(z)\big) &= \frac{\|f - f^*\|}{\|f\|} - \frac{\|z - z^*\|}{\|z\|} \\
&\leq \frac{\|f - f^*\|}{(1 - \delta^2/36)\|z\|} - \frac{\|z - z^*\|}{\|z\|} \\
&\leq \frac{\|f - f^*\|(1 + \delta^2/18) - \|z - z^*\|}{\|z\|} \\
&\leq \frac{(1 + \delta^2/18)(\|z - z^*\| + (\delta/54)\alpha(\ell_{j-1})) - \|z - z^*\|}{\|z\|} \\
&\leq \frac{\delta^2}{18}\frac{\|z - z^*\|}{\|z\|} + \frac{\delta}{54}\frac{\alpha(\ell_{j-1})}{\|z\|} + \frac{\delta^3}{18 * 54}\frac{\alpha(\ell_{j-1})}{\|z\|} \\
&\leq \frac{\delta^2}{18}\alpha(\ell_{j-1}) + \frac{\delta}{18}\alpha(\ell_{j-1}) + \frac{\delta^3}{162}\alpha(\ell_{j-1}) \\
&\leq \frac{\delta}{5}\alpha(\ell_{j-1}).
\end{aligned}
$$

So for any $f \in \overline{ab}$, we have

$$
\begin{aligned}
\alpha\big(\ell(f)\big) &\leq \alpha\big(\ell(z)\big) + \frac{\delta}{5}\alpha(\ell_{j-1}) \leq (1 - \delta/4)\alpha(\ell_{j-1}) + (\delta/5)\alpha(\ell_{j-1}) \\
&\leq (1 - \delta/20)\alpha(\ell_{j-1}).
\end{aligned}
$$

This completes the proof of the claim as the probability that $\alpha(\ell_j) \leq (1 - \delta/20)\alpha(\ell_{j-1})$ is bounded below by the probability of w being a witness times the probability of choosing the point that defines ℓ_j from the segment $\overline{ab}$ given that w was a witness, and this is at least $(\delta/4)^\tau \cdot (\delta^2/216) \cdot (1/2)$. $\qquad\square$

Theorem 1.1 immediately follows from the following Lemma (by invoking `Good-Subspace` with $\mathcal{S} = \mathbb{R}^d$).

Lemma 4.2 *For any inputs P, $\mathcal{S}$, $1 \leq k \leq \dim(\mathcal{S})$, and δ, the algorithm* `Good-Subspace` *returns a k-subspace $\hat{F}_k$ of $\mathcal{S}$ with the guarantee that*

$$\Pr\left(\mathcal{RD}_\tau\big(\hat{F}_k, P\big) \leq (1 + \delta)^k \mathcal{RD}_\tau\big(F^*, P\big)\right) \geq \Delta^k,$$

where

$$\Delta = \frac{\mu^i}{2(i+1)}.$$

Proof The lemma clearly holds if P contains no point different from the origin. So henceforth we assume that this is not the case and prove by induction on k. The base case, $k = 1$, is furnished by Lemma 4.1. For the induction step, assume $k > 1$. By Lemma 4.1 we have that $\mathcal{RD}_\tau(\hat{F}_1, P) \le (1+\delta)\mathcal{RD}_\tau(F^*, P)$ with a probability of at least Δ. Given this event, $\pi(\hat{F}_1)$ is a $(k-1)$-subspace of $\mathcal{S}'$ such that

$$\mathcal{RD}_\tau\big(\pi(\hat{F}_1), \pi(P)\big) = \mathcal{RD}_\tau\big(\hat{F}_1, P\big) \le (1+\delta)\mathcal{RD}_\tau\big(F^*, P\big).$$

Thus by induction hypothesis, the $(k-1)$-subspace G returned by the recursive call satisfies

$$\mathcal{RD}_\tau\big(G, \pi(P)\big) \le (1+\delta)^{k-1}\mathcal{RD}_\tau\big(\pi(\hat{F}_1), \pi(P)\big)$$

with a (conditional) probability of at least Δ^{k-1}. It follows that

$$\mathcal{RD}_\tau\big(\hat{F}_k, P\big) = \mathcal{RD}_\tau\big(G, \pi(P)\big) \le (1+\delta)^{k-1}\mathcal{RD}_\tau\big(\pi(\hat{F}_1), \pi(P)\big)$$
$$\le (1+\delta)^k \mathcal{RD}_\tau\big(F^*, P\big)$$

with a probability of at least Δ^k. □

5 Conclusion

We have shown that for any set P of n points in $\mathbb{R}^d$, and constants $\tau \ge 1$, $0 < \varepsilon < 1$, and integer $k \ge 1$, we can find in time $O(nd)$ a k-subspace F that with probability at least $1/2$ has the property that $\mathcal{RD}_\tau(F, P) \le (1+\varepsilon)\mathcal{RD}_\tau(F', P)$ for any k-subspace F'. The "constant" in the running time is exponential in $\frac{\tau k}{\varepsilon}$. This is the first result of its kind for each $\tau \ne 2, \infty$. A natural question is whether algorithms with similar running times can be obtained when the distance $d(p, F)$ of point p from subpace F (that is used in measuring the fit $\mathcal{RD}_\tau(F, P)$) is defined to be $\min_{x \in F} \|p - x\|_M$ where now $\|\cdot\|_M$ is a norm that is from a larger class of norms that includes the Euclidean norm.

Appendix: Proof of Corollary 1.2

The algorithm in Theorem 1.1 fails with a probability of at most $(1 - \frac{1}{\alpha})$, where $\alpha = 2^{O(\frac{\tau k^2}{\varepsilon} \log^2 \frac{k}{\varepsilon})}$. We run α independent trials of the algorithm in Theorem 1.1 and return the best solution.

Observe that the overall algorithm fails (to return a subspace whose fit is at most $(1 + \varepsilon)$ times the optimal) only if every one of the trials fails. The probability that every one of the trials fails is at most $(1 - \frac{1}{\alpha})^{\alpha} < \frac{1}{e} < \frac{1}{2}$.

The bound on the running time follows from a straight calculation.

References

1. Agarwal, P., Har-Peled, S., Varadarajan, K.: Geometric approximation via coresets. In: Goodman, J., Pach, J., Welzl, E. (eds.) Combinatorial and Computational Geometry. Cambridge University Press, Cambridge (2005)
2. Bădoiu, M., Indyk, P.: Fast approximation algorithms for the hyperplane fitting problem. Manuscript (2006)
3. Bădoiu, M., Har-Peled, S., Indyk, P.: Approximate clustering via core-sets. In: Proc. 34th Annu. ACM Sympos. Theory Comput., pp. 250–257 (2002)
4. Brieden, A.: Geometric optimization problems likely not contained in apx. D C G, **28**(2), 201–209 (2002)
5. Brieden, A., Gritzmann, P., Kannan, R., Klee, V., Lovasz, L., Simonovits, M.: Approximation of diameters: randomization doesn't help. In: Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS), pp. 244–251 (1998)
6. Chakraborty, B., Chaudhuri, P.: A note on the robustness of multivariate medians. Stat. Probab. Lett. **45**, 269–276 (1999)
7. Chandrasekharan, R., Tamir, A.: Algebraic optimization: the Fermat–Weber location problem. Math. Program. **46**, 219–224 (1990)
8. Clarkson, K.: Subgradient and sampling algorithms for L_1 regression. In: Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 257–266 (2005)
9. Deshpande, A., Varadarajan, K.R.: Sampling-based dimension reduction for subspace approximation. In: Proceedings of the 39th ACM Symposium on Theory of Computing (STOC), pp. 641–650 (2007)
10. Deshpande, A., Vempala, S.: Adaptive sampling and fast low-rank matrix approximation. In: Proceedings of APPROX-RANDOM, pp. 292–303 (2006)
11. Deshpande, A., Rademacher, L., Vempala, S., Wang, G.: Matrix approximation and projective clustering via volume sampling. In: Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1117–1126 (2006)
12. Diaz-Banez, J.M., Mesa, J.A., Schobel, A.: Continuous location of dimensional structures. Eur. J. Oper. Res. **152**, 22–44 (2004)
13. Drineas, P., Mahoney, M.W., Muthukrishnan, S.: Sampling algorithms for L_2 regression and applications. In: Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1127–1136 (2005)
14. Faigle, U., Kern, W., Streng, M.: Note on the computational complexity of j-radii of polytopes in $\mathbb{R}^n$. Math. Program. **73**, 1–5 (1996)
15. Feldman, D., Fiat, A., Sharir, M.: Coresets for weighted facilities and their applications. In: Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS), pp. 315–324 (2006)
16. Frieze, A., Kannan, R., Vempala, S.: Fast Monte-Carlo algorithms for finding low rank approximations. J. ACM **51**(6), 1025–1041 (2004)
17. Gritzmann, P., Klee, V.: Computational complexity of inner and outer j-radii of polytopes in finite-dimensional normed spaces. Math. Program. **59**, 163–213 (1993)
18. Gritzmann, P., Klee, V.: On the complexity of some basic problems in computational convexity: I. Containment problems. Discrete Math. **136**, 129–174 (1994)
19. Har-Peled, S.: Low rank matrix approximation in linear time. Manuscript (2006)
20. Har-Peled, S.: How to get close to the median shape. Comput. Geom. **36**(1), 39–51 (2007)
21. Har-Peled, S., Varadarajan, K.R.: Projective clustering in high dimensions using core-sets. In: Proc. 18th Annu. ACM Sympos. Comput. Geom, pp. 312–318 (2002)
22. Har-Peled, S., Varadarajan, K.: High-dimensional shape fitting in linear time. Discrete Comput. Geom. **32**(2), 269–288 (2004)
23. Korneenko, N.M., Martini, H.: Hyperplane approximation and related topics. In: Pach, J. (ed.) New Trends in Discrete and Computational Geometry, pp. 135–162. Springer, New York (1993)

24. Mardia, K.V., Kent, J.T., Bibby, J.M.: Multivariate Analysis. Academic Press, London (1979)
25. Megiddo, N.: On the complexity of some geometric problems in unbounded dimension. J. Symb. Comput. **10**, 327–334 (1990)
26. Panigrahy, R.: Minimum enclosing polytope in high dimensions. arXiv:cs.CG/0407020
27. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
28. Sarlos, T.: Improved approximation algorithms for large matrices via random projections. In: Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS), pp. 143–152 (2006)
29. Tenenbaum, J., de Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
30. Varadarajan, K.R., Venkatesh, S., Zhang, J., Ye, Y.: Approximating the radii of point sets. SIAM J. Comput. **36**(6), 1764–1776 (2007)
31. Zhang, Z., Zha, H.: Principal manifolds and nonlinear dimension reduction via local tangent space alignment. SIAM J. Sci. Comput. **26**(1), 313–338 (2005)