Efficient Subspace Approximation Algorithms

Nariankadu D. Shyamalkumar* Kasturi Varadarajan†

1 Introduction

Motivated by applications in data mining, statistics, and clustering, we consider the problem of fitting a flat of a specified dimension to a finite set P of n points in $\mathbb{R}^d$. A *flat* (resp., k-flat) F in $\mathbb{R}^d$ is defined to be a translation of a subspace (resp., k dimensional subspace). Specifically, we are interested in the following *approximate flat fitting* problem: Given P as above, an integer $0 \leq k \leq d-1$, a measure $\mathcal{RD}(F', P)$ of the fit of any flat F' to P, and a parameter $\varepsilon \geq 0$, find a k-dimensional flat F such that $\mathcal{RD}(F, P) \leq (1+\varepsilon)\mathcal{RD}(F', P)$ for every k-dimensional flat F'. We will refer to the special case where $\varepsilon = 0$ as the *exact* flat fitting problem.

For each $\tau \geq 1$, 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 $d(p, F) = \min_{x \in F} |px|$ is the minimum Euclidean distance between p and a point in F.[1] In this article we will consider the flat fitting problem with such measures, a problem which has received considerable attention, particularly for the cases $\tau = 1, 2, \infty$. Our main result is that for any $\tau \geq 1$, 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 exactly in polynomial time;[2] see for instance [14]. The case $k = 1$, the minimum enclosing cylinder problem, is NP-hard [22]. For any fixed k, there are algorithms that solve the problem in $O(ndC_{\varepsilon,k})$ time, where $C_{\varepsilon,k}$ is a constant that is exponential in $2^k/\varepsilon$ [19, 23]. For large k, the problem becomes hard to approximate in polynomial time to within a factor of $(\log n)^\delta$, for some $\delta > 0$ [4, 25]. The best known polynomial time approximation algorithms yield an approximation guarantee of $O(\sqrt{\log n})$ [26]. If d is a constant, the problem can be solved exactly in polynomial time for every k [16, 11]; ε-approximation algorithms with running time near linear in n 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 flat fitting problem. For instance, it is well known that the optimal k-subspace is obtained by the span of the k right singular vectors corresponding

*University of Iowa; Department of Statistics and Actuarial Science; shyamal-kumar@uiowa.edu

†University of Iowa; Department of Computer Science; kvaradar@cs.uiowa.edu

[1] In the case $\tau = \infty$, $\mathcal{RD}_\tau(F, P)$ is naturally defined as $\max_{p \in P} d(p, F)$.

[2] We will view algorithms that solve the ε-approximate version of the problem for any $0 < \varepsilon < 1$ in time that is polynomial in $\log 1/\varepsilon$ as algorithms that solve the problem exactly.

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, $O(nd \min\{n, d\})$) time algorithm for this problem; see the discussion in [7]. For the ε-approximate problem for small k, recent works give algorithms that are near linear in $\frac{ndk}{\varepsilon}$ [2, 8, 17, 24].

The case $\tau = 1$ and $k = 0$ is the *Fermat-Weber* problem, which reduces to minimizing a convex function over $\mathbb{R}^d$. A polynomial time algorithm for the problem is given by [5]. 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 this problem; see the surveys [20, 9]. 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 [16, 12].

A problem related to the hyperplane problem is the well-studied regression problem; we refer to [6, 10] for some recent work on this.

Results, Techniques, and Related Work

Our main result, stated for the $\tau = 1$ case, is the following.

Theorem 1.1 *There is a randomized algorithm that, given any set P of n points in $\mathbb{R}^d$, any $1 \le k < d$ and any $0 < \varepsilon < 1$, runs in $O(\frac{ndk}{\varepsilon} \log \frac{1}{\varepsilon})$ time and returns with probability $2^{-O(\frac{k}{\varepsilon} \log^2 \frac{1}{\varepsilon})}$ a k-subspace F such that $\mathcal{RD}_1(F, P) \le (1 + \varepsilon)^k \mathcal{RD}_1(F', P)$, for any k-subspace F'.*[3]

The theorem generalizes to the case of $\tau \ge 1$, with the success probability becoming $2^{-O(\frac{\tau k}{\varepsilon} \log^2 \frac{1}{\varepsilon})}$. For ease of presentation, and due to the similarity of the arguments, we restrict ourselves to the case $\tau = 1$.

The randomized algorithm referred to in the theorem works by guessing a sequence of $O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$ lines such that with probability $2^{-O(\frac{1}{\varepsilon} \log^2 \frac{1}{\varepsilon})}$ 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 algorithms returns the k-subspace spanned by ℓ and this $(k-1)$-subspace.

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 [19], 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 [13] 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 [13] and

[3] Of course, one can repeat the algorithm $2^{O(\frac{k}{\varepsilon} \log^2 \frac{1}{\varepsilon})}$ times and take the best solution to increase the success probability to 1/2.

Deshpande et al [7], 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 τ-th 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 $\tau \geq 1$.

A comparison with the results of [8, 17] for the case $\tau = 2$ is useful. These results rely on Theorem 6 from [7] 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 [8, 17, 24] take, enabling a running time that is nd poly $\left(\frac{k}{\varepsilon}\right)$. However, for the case of $\tau = 1$, known ε-approximation algorithms are exponential in the dimension. Hence a reduction of the dimension to $O(k \log n/\varepsilon)$ would not bring us close to the results of Theorem 1.1, and a reduction of the dimension to $O(k/\varepsilon)$ would not eliminate the exponential dependence of the running time on k/ε.

The following theorem extends Theorem 1.1 to the affine case. It follows relatively easily from Theorem 1.1, and we omit its proof from this abstract.

Theorem 1.2 *There is a randomized algorithm that, given any set P of n points in $\mathbb{R}^d$, any $1 \leq k < d$ and any $0 < \varepsilon < 1$, runs in $O(\frac{ndk}{\varepsilon} \log \frac{1}{\varepsilon})$ time and returns with probability $2^{-O(\frac{k}{\varepsilon} \log^2 \frac{1}{\varepsilon})}$ a k-flat F such that $\mathcal{RD}_1(F, P) \leq (1+\varepsilon)^{k+1} \mathcal{RD}_1(F', P)$ for any k-flat F'.*

One consequence of our techniques is the following structural result. Although it is implicit in the proof of Theorem 1.1, we present a separate proof that is not only much simpler, but which also motivates the algorithm of Theorem 1.1.

Theorem 1.3 *Fix $\tau \geq 1$. Let P be a set of points in $\mathbb{R}^d$ and let F^* be the k-dimensional flat that minimizes $\mathcal{RD}_\tau(\cdot, P)$ over all k-dimensional flats. In the nontrivial case of $\mathcal{RD}_\tau(F^*, P) > 0$, and any $0 < \varepsilon < 1$, there exists a $Q \subset P$ consisting of $O(\frac{k}{\varepsilon} \log \frac{1}{\varepsilon})$ points such that the span of Q contains a k-flat F such that $\mathcal{RD}_\tau(F, P) \leq (1+\varepsilon)^{k+1} \mathcal{RD}_\tau(F^*, P)$.*

Similar results for the cases $\tau = \infty$ and $\tau = 2$ were shown by Har-Peled and Varadarajan [18] and Deshpande et al. [7], respectively. Theorem 1.3 has applications to the *projective clustering* problem. For more details we refer to [7, 18].

The remainder of this paper is organized as follows. In Section 2, we describe some useful geometric concepts and tools. In Section 3, we establish Theorem 1.1. For lack of space, the proof of Theorem 1.3 is described in the appendix.

2 Preliminaries

The following extends Lemma 2.5 of [19].

Lemma 2.1 *Let p^*, p, and q be three points in $\mathbb{R}^d$ such that $|qp| \geq (1+\varepsilon)|qp^*|$, where $0 < \varepsilon \leq 1$. (i) Then there exists a point r on the segment $\overline{pq}$ such that $|p^*r| \leq (1-\varepsilon/2)|p^*p|$. (ii) Moreover, if e is chosen uniformly at random from the segment $\overline{pq}$ then $|p^*e| \leq (1-\varepsilon/3)|p^*p|$ with probability at least $\varepsilon/4$.*

Proof: Let $\rho = |qp^*|/|qp|$, and let r be the point on the segment $\overline{pq}$ at a distance $\rho|qp^*|$ from q. It is easy to see that $\triangle qrp^*$ is similar to $\triangle qp^*p$ with a scaling factor of ρ. Therefore,

$$|rp^*| = \rho|p^*p| \leq |p^*p|/(1+\varepsilon) \leq (1-\varepsilon/2)|p^*p|.$$

To show (ii), first note that $|rp|/|pq| = 1-\rho^2 \geq 3\varepsilon/4$. Let e be a point on the segment $\overline{rp}$ and $\alpha \in [0,1]$ be such that $e = \alpha r + (1-\alpha)p$. Then

$$\frac{|p^*e|}{|p^*p|} = \frac{||\alpha(r-p^*) + (1-\alpha)(p-p^*)||}{|p^*p|} \leq \alpha\frac{|rp^*|}{|p^*p|} + (1-\alpha) \leq 1 - \varepsilon\alpha/2,$$

where the last inequality used (i).

Thus, if $\alpha \geq 2/3$, we have $|p^*e| \leq (1-\varepsilon/3)|p^*p|$. Now the probability that $e = \alpha r + (1-\alpha)p$ for $2/3 \leq \alpha \leq 1$ when e is chosen uniformly at random from the segment $\overline{pq}$ is at least $(1-2/3)|rp|/|pq| \geq \varepsilon/4$.

■

Lemma 2.2 *Let F be a k-flat in $\mathbb{R}^d$, p a point not on F, G the translation of F through p, and q a point such that $d(q,G) \geq (1+\varepsilon)d(q,F)$. Then there exists a point r on the segment $\overline{pq}$ such that $d(r,F) \leq (1-\varepsilon/2)d(p,F)$. (ii) Moreover, if e is chosen uniformly at random from the segment $\overline{pq}$ then $d(e,F) \leq (1-\varepsilon/3)d(p,F)$ with probability at least $\varepsilon/4$.*

Proof: Let π be the linear projection onto the orthogonal complement of (the subspace parallel to) F. Any translate F' of F is mapped to a point, that is, $\pi(F')$ is a point. Moreover for any $a \in \mathbb{R}^d$, $d(a,F')$ equals the distance between the points $\pi(F')$ and $\pi(a)$.

Therefore, we have $|\pi(q)\pi(G)| \geq (1+\varepsilon)|\pi(q)\pi(F)|$. From Lemma 2.1 (i), there exists a point r' on the segment $\overline{\pi(G)\pi(q)}$ such that $|\pi(F)r'| \leq (1-\varepsilon/2)|\pi(F)\pi(G)|$. We have $d(p,F) = |\pi(G)\pi(F)|$. Since $\pi(G) = \pi(p)$, the segment $\overline{\pi(G)\pi(q)}$ is the same as the segment $\overline{\pi(p)\pi(q)}$. Suppose $r' = \alpha\pi(p) + (1-\alpha)\pi(q)$ for some $0 \leq \alpha \leq 1$. Then $r' = \pi(r)$, where $r = \alpha p + (1-\alpha)q$ is a point on segment $\overline{pq}$. We have $|\pi(F)r'| = d(r,F)$. Thus, $d(r,F) \leq (1-\varepsilon/2)d(p,F)$.

To prove part (ii), we proceed as above using Lemma 2.1 (ii), and the observation that picking a point uniformly at random from the segment $\overline{pq}$ is the same as picking a point e' uniformly at random from $\overline{\pi(p)\pi(q)}$ and returning the point e such that $\pi(e) = e'$.

■

The following lemma is in the same spirit as Lemma 2.4 of [19].

Lemma 2.3 *Let d_G be the distance function to a flat G. Let $x, y \in \mathbb{R}^d$ be any two points, and let w, z be any two points on the line through x and y. Then*

$$\frac{|d_G(z) - d_G(w)|}{|zw|} \leq \frac{d_G(x) + d_G(y)}{|xy|}.$$

Proof: We assume without loss of generality that G is a subspace. 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 $d_G(a) = ||a_{G^\perp}||$. We have

$$\begin{aligned}\frac{|d_G(z) - d_G(w)|}{|zw|} &\leq \frac{||z_{G^\perp} - w_{G^\perp}||}{|zw|} = \frac{||x_{G^\perp} - y_{G^\perp}||}{|xy|} \\ &\leq \frac{||x_{G^\perp}|| + ||y_{G^\perp}||}{|xy|} = \frac{d_G(x) + d_G(y)}{|xy|}.\end{aligned}$$

∎

We conclude this section by defining notion of the rotation of a k-subspace F through a line ℓ that passes through the origin. If the projection of ℓ onto F is the origin, then we take any $(k-1)$-subspace H of F, and define the rotation to be the k-subspace spanned by H and ℓ. Otherwise, the projection of ℓ onto F is a line ℓ'. We take H to be the orthogonal complement of ℓ' in F, and the rotation to be the k-subspace spanned by H and ℓ.

3 Efficient Computation of Good Subspaces

In this section, we describe the algorithm and the analysis needed to establish Theorem 1.1. Throughout this section, we use $\mathcal{RD}(\cdot,\cdot)$ to mean $\mathcal{RD}_1(\cdot,\cdot)$.

The Algorithm

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 S, an integer $1 \leq k < \dim(\mathcal{S})$, and a parameter $0 < \varepsilon < 1$. It returns a k-dimensional subspace F of $\mathcal{S}$, and we will later argue that $\mathcal{RD}(F, P) \leq (1+\varepsilon)^k \mathcal{RD}(F', P)$ for any k-dimensional subspace F' of $\mathcal{S}$ with a reasonably large probability. The parameter on which the algorithm recurses is k; the base case will be $k = 1$.

If every point in the multiset P is the same as o, we return any k-subspace lying in $\mathcal{S}$. Otherwise, we first compute a sequence $\ell_0, \ldots, \ell_i$ of lines, where $i = \lceil \frac{c}{\varepsilon} \log \frac{1}{\varepsilon} \rceil$ 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 $\frac{|oq|}{\sum_{p \in P|} |op|}$ and set $\ell_0 = \ell(p)$. 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 r from P according to the same distribution used above. Let u and v be unit vectors in the direction ℓ_j and $\ell(r)$, 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 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 ε. 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 returns.

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 $O(\frac{nd}{\varepsilon}\log\frac{1}{\varepsilon})$. Since the depth of the recursion is k, the overall running time of the algorithm is $O(\frac{ndk}{\varepsilon}\log\frac{1}{\varepsilon})$.

Performance

Let F^* denote the k-subspace in S that minimizes $\mathcal{RD}(\cdot, P)$. Let F_j denote the rotation of F^* through ℓ_j, for $1 \leq j \leq i$, and F the rotation of F^* through ℓ. For a point $p \neq o$, we let $\ell(p)$ denote the line through o and p. The following lemma is the essence of the performance guarantee of the algorithm.

Lemma 3.1 *Suppose that P contains some point that is different from o. With a probability of at least $(\varepsilon^3/1728)^i/2$, there exists a j between 0 and i such that $\mathcal{RD}(F_j, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$. Consequently, with a probability of at least $(\varepsilon^3/1728)^i/2(i+1)$, we have $\mathcal{RD}(F, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$.*

Proof: For any $p \in \mathbb{R}^d$, let $\bar{p}$ 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, $\alpha(\ell) = |p\bar{p}|/|op|$ for any point $p \neq 0$ on ℓ. A calculation shows that

$$(\sum_{p\in P}|op|)E[\alpha(\ell_0)] = \mathcal{RD}(F^*, P).$$

Using Markov's inequality, we conclude that with a probability of at least 1/2, we have

$$(\sum_{p\in P}|op|)\alpha(\ell_0) \leq 2\mathcal{RD}(F^*, P) \tag{1}$$

We also need the following claim, whose proof we describe after showing how it implies the lemma.

Claim: For any $1 \leq j \leq i$, suppose $\ell_0, \ldots, \ell_{j-1}$ are such that ℓ_0 satisfies the inequality (1), $\alpha(\ell_{j'}) \leq \alpha(\ell_{j'-1})$ for $1 \leq j' \leq j-1$, and $\mathcal{RD}(F_{j-1}, P) > (1+\varepsilon)\mathcal{RD}(F^*, P)$. Then the probability that $\alpha(\ell_j) \leq (1-\varepsilon/20)\alpha(\ell_{j-1})$, given such $\ell_0, \ldots, \ell_{j-1}$, is at least $\varepsilon^3/1728$.

Assuming the claim, it follows that with a probability of at least $(\varepsilon^3/1728)^i/2$, the following events simultaneously occur:

1. $(\sum_{p\in P}|op|)\alpha(\ell_0) \leq 2\mathcal{RD}(F^*, P)$, and
2. For $1 \leq j \leq i$, either $\mathcal{RD}(F_{j-1}, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$ or $\alpha(\ell_j) \leq (1-\varepsilon/20)\alpha(\ell_{j-1})$.

We argue that these events imply that $\mathcal{RD}(F_j, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$ 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

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

by our choice of i. Denoting by p' the projection of $\bar{p}$ onto F_i, we have

$$\begin{aligned} \mathcal{RD}(F_i, P) &\leq \sum_{p \in P} (|p\bar{p}| + |\bar{p}p'|) \leq \mathcal{RD}(F^*, P) + \sum_{p \in P} |o\bar{p}| \alpha(\ell_i) \\ &\leq \mathcal{RD}(F^*, P) + \frac{\varepsilon}{2} \alpha(\ell_0) \sum_{p \in P} |op| \leq (1+\varepsilon)\mathcal{RD}(F^*, P). \end{aligned}$$

Proof of Claim: Let us call a point $p \in P$ a *witness* if $d(p, F_{j-1}) > (1+\varepsilon/2)d(p, F^*)$ and let P_{j-1} be the set of all witnesses. We claim that

$$\sum_{p \in P_{j-1}} |op| \geq \frac{\varepsilon}{4} \sum_{p \in P} |op|. \tag{2}$$

If this does not hold, then

$$\begin{aligned} \sum_{p \in P} d(p, F_{j-1}) &\leq (1+\varepsilon/2) \sum_{p \in P \setminus P_{j-1}} d(p, F^*) + \sum_{p \in P_{j-1}} (|p\bar{p}| + d(\bar{p}, F_{j-1})) \\ &\leq (1+\varepsilon/2) \sum_{p \in P} d(p, F^*) + \sum_{p \in P_{j-1}} |o\bar{p}| \alpha(\ell_{j-1}) \\ &\leq (1+\varepsilon/2) \sum_{p \in P} d(p, F^*) + \alpha(\ell_0) \sum_{p \in P_{j-1}} |o\bar{p}| \\ &\leq (1+\varepsilon/2) \sum_{p \in P} d(p, F^*) + \frac{\varepsilon}{4} \alpha(\ell_0) \sum_{p \in P} |o\bar{p}| \\ &\leq (1+\varepsilon/2) \sum_{p \in P} d(p, F^*) + (\varepsilon/2)\mathcal{RD}(F^*, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P), \end{aligned}$$

where the penultimate inequality used the fact that ℓ_0 satisfies inequality (1). But we have arrived at a contradiction to the assumption that $\mathcal{RD}(F_{j-1}, P) > (1+\varepsilon)\mathcal{RD}(F^*, P)$.

The inequality (2) means that the point r chosen by the algorithm in constructing ℓ_j from ℓ_{j-1} has a probability of at least $\varepsilon/4$ of being a witness. Let us assume that this event happens, that is, let us condition on it.

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

Let $\hat{r} = \pi_H(r) = \pi_H(\bar{r}) = \pi_H(r')$. We verify that the point $q' = r' - \hat{r}$ lies on the line ℓ_{j-1}. Considering $\triangle rr'q'$, and recalling that s lies on $\overline{r'r}$, we see that there is a point q on the segment $\overline{q'r}$ such that $q - s$ is a scaling of $-\hat{r}$. (If $\hat{r} = o$, q' and q degenerate to r' and s respectively.) Let e be the point on the line $\{\bar{r} - t\hat{r} | t \in \mathbb{R}\}$ closest to q. (If $\hat{r} = o$, then $e = \bar{r}$.) It is easy to verify that $|eq| \leq |\bar{r}s|$ since $\bar{r}$ and s are on lines parallel to $-\hat{r}$ and $|eq|$ is the distance between these lines. Finally, let e' be the projection of e onto F_{j-1}. Since e

is a translation of $\bar{r}$ by a vector that is scale of $-\hat{r}$ and which therefore lies in F_{j-1}, we have $|\bar{r}r'| = |ee'|$. So we have

$$|eq| \leq |\bar{r}s| \leq (1-\varepsilon/4)|\bar{r}r'| = (1-\varepsilon/4)|ee'|.$$

To bound $\alpha(\ell(q))$, it is enough to bound the sine of the angle between $\ell(q)$ and $\ell(e)$, since e lies on F^*. Thus

$$\alpha(\ell(q)) \leq \frac{|eq|}{|oe|} \leq (1-\varepsilon/4)\frac{|ee'|}{|oe|} \leq (1-\varepsilon/4)\alpha(\ell_{j-1}), \tag{3}$$

where the last inequality can be seen from the facts that e lies on F^*, e' is the projection of e onto F_{j-1}, and F_{j-1} is the rotation of F^* through ℓ_{j-1}.

We have so far shown that $\ell(q)$ is a line lying on the span of ℓ_{j-1} and the sampled point r, and $\ell(q)$ makes a significantly smaller angle with F^* than ℓ_{j-1}. Our next step is to 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 $\frac{r}{|or|}$. Suppose that the inner product $r \cdot u > 0$, and that in fact the angle uor is at most $\pi/4$. In this case, we argue that q' lies on the ray $\{tu | t > 0\}$. First, observe that $\bar{r} \cdot u > 0$, because if this is not the case we will have $d(r, F^*) = |r\bar{r}| \geq d(r, \ell_{j-1}) \geq d(r, F_{j-1})$, contradicting the fact that r is a witness. Now since the vectors $r' - \bar{r}$ and $q' - r'$ are orthogonal to ℓ_{j-1}, we have $q' \cdot u > 0$, that is, q' lies on the ray $\{tu | t > 0\}$.

Let $w = u / -u$ be the unit vector such that q' lies on the ray $\{tw | t \geq 0\}$. We have just argued that the angle wov is at most $3\pi/4$. Since q lies on the segment $\overline{q'r}$, the line $\ell(q)$ intersects the segment $\overline{wv}$ at some point, call it z. There is a segment $\overline{ab}$ containing z and contained in $\overline{wv}$ so that $\frac{|ab|}{|wv|} = \frac{\varepsilon^2}{216}$. We show that for any point f in this segment, $\alpha(\ell(f)) \leq (1-\varepsilon/20)\alpha(\ell_{j-1})$. This will finish the proof of the claim, because the probability $\alpha(\ell_j) \leq (1-\varepsilon/20)\alpha(\ell_{j-1})$ is bounded below by the probability of choosing a witness r times the probability of choosing the point that defines ℓ_j from the segment $\overline{ab}$ given that r was a witness, and this is at least $(\varepsilon/4) * (\varepsilon^2/216) * (1/2)$.

Since r is a witness, $\frac{\varepsilon}{2}|or|\alpha(\ell(r)) = \frac{\varepsilon}{2}|r\bar{r}| \leq |\bar{r}r'| \leq |o\bar{r}|\alpha(\ell_{j-1}) \leq |or|\alpha(\ell_{j-1})$, so $\alpha(\ell(r)) \leq \frac{2}{\varepsilon}\alpha(\ell_{j-1})$. Using Lemma 2.3, we see that if f is a point on the segment $\overline{ab}$, then

$$|f\bar{f}| - |z\bar{z}| \leq \frac{|ab|}{|vw|}(|v\bar{v}| + |w\bar{w}|) \leq \frac{\varepsilon^2}{216}\frac{4}{\varepsilon}\alpha(\ell_{j-1}) = \frac{\varepsilon}{54}\alpha(\ell_{j-1}).$$

Observe that $|ab| \leq \frac{\varepsilon^2}{216}|vw| \leq \frac{\varepsilon^2}{108}$. Also, for any point f on $\overline{wv}$, we have $|of| \geq 1/3$ as a consequence of the angle vow being at most $3\pi/4$. Thus for any $f \in \overline{ab}$, we have

$$|of| \geq |oz| - |fz| \geq |oz| - \frac{\varepsilon^2}{108} \geq |oz|(1 - \frac{\varepsilon^2}{36}).$$

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

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

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

$$\alpha(\ell(f)) \leq \alpha(\ell(z)) + \frac{\varepsilon}{5}\alpha(\ell_{j-1}) \leq (1-\varepsilon/4)\alpha(\ell_{j-1}) + (\varepsilon/5)\alpha(\ell_{j-1}) \leq (1-\varepsilon/20)\alpha(\ell_{j-1}).$$

■

Lemma 3.2 *For any inputs P, $\mathcal{S}$, $1 \leq k \leq dim(S)$, and $0 < \varepsilon < 1$, the algorithm* `Good-Subspace` *returns with probability at least δ^k, where $\delta = ((\varepsilon^3/1728)^i/2(i+1))$ a k-subspace F' of $\mathcal{S}$ such that $\mathcal{RD}(F', P) \leq (1+\varepsilon)^k \mathcal{RD}(F^*, P)$.*

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 3.1. For the induction step, assume $k > 1$. By Lemma 3.1 we have that $\mathcal{RD}(F, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$ with a probability of at least δ. Given this event, $\pi(F)$ is a $(k-1)$-subspace of $\mathcal{S}'$ such that $\mathcal{RD}(\pi(F), \pi(P)) = \mathcal{RD}(F, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$. Thus by induction hypothesis, the $(k-1)$-subspace G returned by the recursive call satisfies $\mathcal{RD}(G, \pi(P)) \leq (1+\varepsilon)^k \mathcal{RD}(F^*, P)$ with a (conditional) probability of at least δ^{k-1}. It follows that $\mathcal{RD}(F', P) = \mathcal{RD}(G, \pi(P)) \leq (1+\varepsilon)^k \mathcal{RD}(F^*, P)$ with a probability of at least δ^k. ■

Theorem 1.1 immediately follows from this Lemma.

References

[1] P. Agarwal, S. Har-Peled, and K. Varadarajan. Geometric approximation via coresets. in J. Goodman, J. Pach, and E. Welzl (Eds), *Combinatorial and Computational Geometry*, Cambridge University Press, 2005.

[2] M. Bădoiu and P. Indyk. Fast approximation algorithms for the hyperplane fitting problem. Manuscript, 2006.

[3] M. Bădoiu, S. Har-Peled, and P. Indyk. Approximate clustering via core-sets. In *Proc. 34th Annu. ACM Sympos. Theory Comput.*, pages 250–257, 2002.

[4] A. Brieden. Geometric optimization problems likely not contained in apx. *DCG*, 28(2):201–209, 2002.

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

[6] K. Clarkson. Subgradient and sampling algorithms for L_1 regression. *In Proceedings of SODA 2005.*

[7] A. Deshpande, L. Rademacher, S. Vempala, and G. Wang. Matrix approximation and projective clustering via volume sampling. In *Proceedings of SODA 2006*.

[8] A. Deshpande and S. Vempala. Adaptive sampling and fast low-rank matrix approximation. To appear in *Proceedings of RANDOM 2006.*

[9] J. M. Diaz-Banez, J. A. Mesa, and A. Schobel Continuous location of dimensional structures. *European Journal of Operations Research*, 152:22–44, 2004.

[10] P. Drineas, M. .W. Mahoney, and S. Muthukrishnan. Sampling algorithms for L_2 regression and applications. *In Proceedings of SODA 2005*

[11] U. Faigle, W. Kern, and M. Streng. Note on the computational complexity of j-radii of polytopes in $\mathbb{R}^n$. *Mathematical Programming*, 73:1–5, 1996.

[12] D. Feldman, A. Fiat, and M. Sharir. Coresets for weighted facilities and their applications. To appear in FOCS 06.

[13] A. Frieze, R. Kannan, and S. Vempala Fast Monte-Carlo algorithms for finding low rank approximations. *J. ACM*, 51(6):1025–1041, 2004.

[14] P. Gritzmann and V. Klee. Computational complexity of inner and outer j-radii of polytopes in finite-dimensional normed spaces. *Math. Program.*, 59:163–213, 1993.

[15] P. Gritzmann and V. Klee. On the complexity of some basic problems in computational convexity: I. containment problems. *Discrete Math.*, 136:129–174, 1994.

[16] S. Har-Peled. How to get close to the median shape. In *Proc. ACM Symp. Comput. Geom.*, 2006.

[17] S. Har-Peled. Low rank matrix approximation in linear time. Manuscript, 2006.

[18] S. Har-Peled and K. R. Varadarajan. Projective clustering in high dimensions using core-sets. In *Proc. 18th Annu. ACM Sympos. Comput. Geom.*, pages 312–318, 2002.

[19] S. Har-Peled and K. Varadarajan, "High-Dimensional Shape Fitting in Linear Time," *Discrete & Computational Geometry*, 32(2): 269-288, 2004.

[20] N. M. Korneenko and H. Martini Hyperplane approximation and related topics. In J. Pach (Ed.), New Trends in Discrete and Computational Geometry, Springer-Verlag, New York, 1993, pp. 135–162.

[21] H. Martini. Minsum k-flats of finite point sets in $\mathbb{R}^d$. *Studies in Locational Analysis*, 7:123–129, 1994.

[22] N. Megiddo. On the complexity of some geometric problems in unbounded dimension. *J. Symb. Comput.*, 10:327–334, 1990.

[23] R. Panigrahy. Minimum enclosing polytope in high dimensions. In arXiv.org, cs.CG/0407020.

[24] T. Sarlos. Improved Approximation algorithms for large matrices via random projections. To appear in *FOCS 2006*.

[25] K. R. Varadarajan, S. Venkatesh, and J. Zhang. Approximating the radii of point sets in high dimensions. In *Proc. 43th Annu. IEEE Sympos. Found. Comput. Sci.*, 2002.

[26] Y. Ye and J. Zhang. An improved algorithm for approximating the radii of point sets. In *Proc. APPROX*, 2003.

Appendix: Existence of Good, Anchored k-Flats

In this section, we prove Theorem 1.3. Let us fix any $\tau \geq 1$ and let $\mathcal{RD}(\cdot,\cdot)$ stand for $\mathcal{RD}_\tau(\cdot,\cdot)$ throughout this section.

Finding a Point on a Nearly Optimal Flat

Lemma 3.3 *Let F^* be the k-flat that minimizes $\mathcal{RD}(F,P)$ for all k-flats F. For any $0 < \varepsilon < 1$, there exists a set of $O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ points of P whose convex hull contains a point q such that the translation F^*_q of F^* through q satisfies $\mathcal{RD}(F^*_q,P) \leq (1+\varepsilon)\mathcal{RD}(F^*,P)$.*

Proof: We construct a sequence $q_0,\ldots,q_i$ of points, 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 closest to F^*.
2. q_j is in the convex hull of at most $j+1$ points from P.
3. Let F_j denote the translate of F^* through q_j. For $1 \leq j \leq i$, if $\mathcal{RD}(F_{j-1},P) > (1+\varepsilon)\mathcal{RD}(F^*,P)$ then $d(q_j,F^*) \leq (1-\varepsilon/2)d(q_{j-1},F^*)$.

The sequence consisting of just the one point q_0 clearly satisfies the last two conditions. Supppose that we have inductively constructed the sequence $q_0,\ldots,q_{j-1}$, where $j \geq 1$. We descibe how to extend the sequence. If $\mathcal{RD}(F_{j-1},P) \leq (1+\varepsilon)\mathcal{RD}(F^*,P)$ this is trivial because we can take q_j to be any point of P. If $\mathcal{RD}(F_{j-1},P) > (1+\varepsilon)\mathcal{RD}(F^*,P)$, there exists a point $q \in P$ such that $d(q,F_{j-1}) > (1+\varepsilon)d(q,F^*)$. Applying Lemma 2.2, there exists a point r on the segment $\overline{q_{j-1}q}$ such that $d(r,F^*) \leq (1-\varepsilon/2)d(q_{j-1},F^*)$. We set q_j to be r. The extended sequence clearly satisfies condition (3). Since q_{j-1} is in the convex hull of at most j points of P, and q_j is in the convex hull of q_{j-1} and q, condition (2) also follows.

We now argue that at least one of the flats F_j satisfies $\mathcal{RD}(F_j,P) \leq (1+\varepsilon)\mathcal{RD}(F^*,P)$, thus proving the lemma. If $F_0,\ldots,F_{i-1}$ do not satisfy this requirement then condition (3) tells us that $d(q_k,F^*) \leq (1-\varepsilon/2)^i d(q_0,F^*) \leq \varepsilon d(q_0,F^*)$ by our choice of i. Then for any $p \in P$, we have $d(p,F_i) \leq d(p,F^*) + d(q_k,F^*) \leq d(p,F^*) + \varepsilon d(q_0,F^*) \leq (1+\varepsilon)d(p,F^*)$, where the last inequality follows from the choice of q_0. But then it follows that $\mathcal{RD}(F_i,P) \leq (1+\varepsilon)\mathcal{RD}(F^*,P)$.

■

Anchoring a Nearly Optimal Line

Lemma 3.4 *Let P be a set of points in $\mathbb{R}^d$, and let ℓ^* be the line that minimizes $\mathcal{RD}(\ell,P)$ over all lines ℓ through the origin o. Assume that $\mathcal{RD}(\ell^*,P) > 0$. 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 $\mathcal{RD}(\ell(t),P) \leq (1+\varepsilon)\mathcal{RD}(\ell^*,P)$, where $\ell(p)$ denotes the line through p (assumed different from o) and o.*

Proof: For any $p \in \mathbb{R}^d$, let $\bar{p}$ denotes its projection onto ℓ^*. For any line ℓ through the origin, let $\alpha(\ell)$ denote the distance $|x\bar{x}|$, 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 $\mathcal{RD}(\ell(q_{j-1}), P) > (1+\varepsilon)\mathcal{RD}(\ell^*, P)$ 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 descibe how to extend the sequence. If $\mathcal{RD}(\ell(q_{j-1}), P) \leq (1 + \varepsilon)\mathcal{RD}(\ell^*, P)$, this is trivial because we can take q_j to be any point of P different from o. Otherwise, there is a point $q \in P$ such that $d(q, \ell(q_{j-1})) > (1 + \varepsilon)d(q, \ell^*)$. Let q' denote the projection of $\bar{q}$ onto $\ell(q_{j-1})$. We have $|qq'| > (1 + \varepsilon)|q\bar{q}|$. From Lemma 2.1, there is a point r on the segment $\overline{q'q}$ such that $|\bar{q}r| \leq (1 - \varepsilon/2)|\bar{q}q'|$. Let $q_j = r$. Since $|\bar{q}q_j| < |\bar{q}q'| \leq |\bar{q}o|$, q_j is different from o. We have

$$d(\bar{q}, \ell(q_j)) \leq |\bar{q}r| \leq (1 - \varepsilon/2)|\bar{q}q'| = d(\bar{q}, \ell(q_{j-1})).$$

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 q, 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 lines $\ell(q_j)$ satisfies $\mathcal{RD}(\ell(q_j), P) \leq (1+\varepsilon)\mathcal{RD}(\ell^*, P)$, thus proving the lemma. If $\ell(q_0), \ldots, \ell(q_{i-1})$ do not satisfy this requirement then condition (3) tells us that $\alpha(\ell(q_i)) \leq (1 - \varepsilon/2)^i\alpha(\ell(q_0)) \leq \varepsilon\alpha(\ell(q_0))$ by our choice of i. Then for any $p \in P$, we have

$$d(p, \ell(q_i)) \leq |p\bar{p}|+|o\bar{p}|\alpha(\ell(q_i)) \leq |p\bar{p}|+|op|\varepsilon\alpha(\ell(q_0)) \leq |p\bar{p}|+\varepsilon|op|\frac{|p\bar{p}|}{|op|} = (1+\varepsilon)|p\bar{p}| = (1+\varepsilon)d(p, \ell^*),$$

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

Anchoring a Nearly Optimal k-Subspace

Lemma 3.5 *Let P be a set of points in $\mathbb{R}^d$ and let F^* be the k-subspace that minimizes $\mathcal{RD}(F, P)$ over all k-subspaces F, where $k \geq 1$. Assume that $\mathcal{RD}(F^*, P) > 0$. For any $0 < \varepsilon < 1$, there exists a set of at most $\frac{ck}{\varepsilon} \log \frac{1}{\varepsilon}$ points in P whose span contains a k-subspace G such that $\mathcal{RD}(G, P) \leq (1+\varepsilon)^k\mathcal{RD}(F^*, P)$. Here $c > 0$ is the constant appearing in Lemma 3.4.*

Proof: The proof is by induction on k. The base case of $k = 1$ is furnished by Lemma 3.4. 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 key 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 3.4, there exists a set $Q_1 \subseteq P$ of at most $\frac{c}{\varepsilon} \log \frac{1}{\varepsilon}$ points such that the span of $\pi(Q_1)$ contains a line ℓ such that $\mathcal{RD}(\ell, \pi(P)) \leq (1+\varepsilon)\mathcal{RD}(\pi(F^*), \pi(P)$. Let F be the k-flat in $\mathbb{R}^d$ spanned by ℓ and $e_1, \ldots, e_{k-1}$. The key property of π implies that $\mathcal{RD}(F, P) \leq (1+\varepsilon)\mathcal{RD}(F^*, P)$. 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 the span of Q_1.

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

$$\mathcal{RD}(G, P) \leq (1+\varepsilon)^{k-1}\mathcal{RD}(\pi'(F), \pi'(P)) = (1+\varepsilon)^{k-1}\mathcal{RD}(F, P) \leq (1+\varepsilon)^{k}\mathcal{RD}(F^*, P).$$

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, $|Q| \leq |Q_1| + |Q_2| \leq \frac{ck}{\varepsilon} \log \frac{1}{\varepsilon}$.

■

Proof of Theorem 1.3: The proof follows by combining Lemma 3.3 and Lemma 3.5.