# Supplementary File for "Randomized Dimensionality Reduction for Facility Location and Single-Linkage Clustering"

**Shyam Narayanan** [*1] **Sandeep Silwal** [*1] **Piotr Indyk** [1] **Or Zamir** [2]

## A. Omitted Preliminaries

In this section, we state all of the preliminary results needed relating to random projections that were omitted in Section 2 of the main paper. In all of the following results, we treat $G$ as a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$.

First, if $x \in S^{m-1}$ then the following statements hold about the distribution of $\|Gx\|$ (Indyk & Naor, 2007):

$$\Pr(|\|Gx\| - 1| \geq t) \leq \exp(-dt^2/8), \tag{1}$$

$$\Pr(\|Gx\| \leq 1/t) \leq \left(\frac{3}{t}\right)^d. \tag{2}$$

The following serves as a converse to Equation (2).

**Proposition A.1.** *If $x \in S^{m-1}$ and $t \geq 1$ then the following is true about the distribution of $\|Gx\|$:*

$$\Pr(\|Gx\| \leq 1/t) \geq \left(\frac{1}{et}\right)^d. \tag{3}$$

*Proof.* Since $x \in S^{m-1}$, $d \cdot \|Gx\|^2$ is a chi-squared random variable with $d$ degrees of freedom so it has density

$$\frac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1} \exp(-x/2).$$

Thus, for all $t \leq 1$, the probability that $\|Ga\|^2$ is less than $1/t^2$ is at least

$$\begin{aligned}
&\frac{1}{2^{d/2}\Gamma(d/2)} \int_0^{d/t^2} x^{d/2-1} \exp(-x/2) dx \\
&\geq \frac{\exp(-d/2t^2)}{2^{d/2}\Gamma(d/2)} \int_0^{d/t^2} x^{d/2-1} dx \\
&\geq \frac{\exp(-d)}{2^{d/2}\Gamma(d/2)(d/2)} \cdot \left(\frac{d}{t^2}\right)^{d/2} \\
&\geq \frac{\exp(-d)}{2^{d/2} \cdot (d/2)^{d/2}} \cdot \left(\frac{d}{t^2}\right)^{d/2} \\
&= \left(\frac{1}{e \cdot t}\right)^d,
\end{aligned}$$

where we used the well-known fact that $\Gamma(x) \cdot x \leq x^x$ for all $x \geq 1$. $\square$

We will need the following lemma to prove some of our lower bound results from Section 6 of the main paper.

**Lemma A.2.** *Let $C \geq 1$ and fix some point $v$ of norm at most $C$ in $\mathbb{R}^d$. Then, if $x \sim \frac{1}{\sqrt{d}} \cdot \mathcal{N}(0, I_d)$ is a $d$-dimensional scaled multivariate Normal, then $\Pr(\|x - v\| \leq \frac{1}{C}) \geq n^{-1/10}$, if $d \leq \log n/(10C^2)$ and $n$ is sufficiently large.*

*Proof.* By the rotational symmetry of the multivariate normal, assume $v = (r, 0, \dots, 0)$, where $0 \leq r \leq C$. Then, if $x = (x_1, y)$ for $x_1 \in \mathbb{R}, y \in \mathbb{R}^{d-1}$, then if $r - \frac{1}{2C} \leq x_1 \leq r$ and $\|y\| \leq \frac{1}{2C}$, then we indeed have $\|x - v\| \leq \frac{1}{C}$. Since $\sqrt{d}x_1 \sim \mathcal{N}(0,1)$ and $r \leq C$, the probability that $r - \frac{1}{2C} \leq x_1 \leq r$ equals the probability that $\mathcal{N}(0,1) \in [(r - 1/2C)\sqrt{d}, r\sqrt{d}]$, which is at least $\frac{\sqrt{d}}{2C} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-C^2 d/2}$. Moreover, the probability that $\|y\| \leq \frac{1}{2C}$ is at least $\left(\frac{1}{2eC}\right)^d$ by Proposition A.1. Therefore,

$$\begin{aligned}
\Pr\left(\|x - v\| \leq \frac{1}{C}\right) &\geq \frac{\sqrt{d}}{2C} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-C^2 d/2} \cdot \left(\frac{1}{2eC}\right)^d \\
&\geq n^{-1/10},
\end{aligned}$$

where the last inequality is true because $d \leq \log n/(10C^2)$ and that $n$ is sufficiently large. $\square$

[*]Equal contribution [1]Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA, USA [2]Institute of Advanced Study, Princeton, NJ, USA. Correspondence to: Shyam Narayanan <shyamsn@mit.edu>, Sandeep Silwal <silwal@mit.edu>, Piotr Indyk <indyk@mit.edu>, Or Zamir <orzamir@ias.edu>.

*Proceedings of the $38^{th}$ International Conference on Machine Learning*, PMLR 139, 2021. 

The following lemma due to Indyk and Naor (2007) was Lemma 2.2 in the main paper, but we restate it here for convenience.

**Lemma A.3** (Lemma 4.2 in (Indyk & Naor, 2007))**.** *Let $X \subseteq B(0,1)$ be a subset of the $m$-dimensional Euclidean unit ball. Then there exist universal constants $c, C > 0$ such that for $d \geq C \cdot d_X + 1$ and $t > 2$, $\Pr(\exists x \in X, \|Gx\| \geq t) \leq \exp(-cdt^2)$.*

Indyk and Naor also prove the following result about the distance to the nearest neighbor after a random projection.

**Theorem A.4** (Theorem 4.1 in (Indyk & Naor, 2007))**.** *Let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(d_X \cdot \log(1/\epsilon)/\epsilon^2 \log(1/\delta))$. Then for every $x \in X$, with probability at least $1 - \delta$, the following statements hold:*

1. *$D(Gx, G(X \setminus \{x\})) \leq (1+\epsilon) D(x, X \setminus \{x\})$*
2. *Every $y \in X$ with $\|x - y\| > (1 + 2\epsilon) D(x, X \setminus \{x\})$ satisfies $\|Gx - Gy\| > (1+\epsilon) D(x, X \setminus \{x\})$ where $D(x, X) = \min_{y \in X} \|x - y\|$.*

# B. The Mettu-Plaxton (MP) Algorithm and Local Optimality

First, we give the pseudocode for the Mettu-Plaxton (MP) algorithm for facility location, described in Section 3 of the main paper.

**Algorithm 1** MP ALGORITHM

**Input :** Dataset $X = \{p_1, \cdots, p_n\} \subseteq \mathbb{R}^d$
**Output :** Set $\mathcal{F}$ of facilities

$\mathcal{F} \leftarrow \emptyset$ **for** $i = 1$ *to* $n$ **do**
  Compute $r_i$ satisfying: $\sum_{q \in B(p_i, r_i)} (r_i - \|p_i - q\|) = 1$
Sort such that $r_1 \leq \ldots \leq r_n$ **for** $i = 1$ *to* $n$ **do**
  **if** $B(p_i, 2r_i) \cap \mathcal{F} = \emptyset$ **then**
    $\mathcal{F} \leftarrow \mathcal{F} \cup \{p_i\}$
Output $\mathcal{F}$

Next, we prove Lemma 3.3 in the main paper, which roughly stated that a globally optimal solution for facility location is always locally optimal.

*Proof of Lemma 3.3 in the main paper.* Consider an arbitrary point $p \in X$. We first establish a lower bound on the number of points in $B(p, r_p) \cap X$. Note that by definition of $r_p$, we have

$$|B(p, r_p) \cap X| r_p \geq \sum_{q \in B(p, r_p)} (r_p - \|p - q\|) = 1$$

so it follows that

$$|B(p, r_p) \cap X| \geq 1/r_p. \tag{4}$$

Now suppose that $B(p, 3r_p) \cap \mathcal{F} = \emptyset$ and let $m$ be the number of points in $|B(p, r_p) \cap X|$ *excluding* $p$. The total connection cost of all these points to their nearest facility must be at least $2mr_p$. Accounting for point $p$, the total connection costs of points in $B(p, r_p) \cap X$ is at least $2mr_p + 3r_p$. Now if we open a new facility at point $p$, then the connection costs of these points is at most $mr_p$ but we also incur an additional cost for opening a facility at $p$. Therefore, the total cost of the solution decreases by at least

$$(2mr_p + 3r_p) - (1 + mr_p) = (m+3)r_p - 1.$$

Now from Eq. (4), we have that $(m+3)r_p > 1$, which means that the total cost decreases if we open a new facility at $p$. $\square$

# C. Dimension Reduction for Facility Location: Omitted Proofs

## C.1. Approximating the Optimal Facility Location Cost

In this subsection, we prove Theorem 4.1 in the main paper. As stated in Subsection 4.1 in the main paper, our proof involves computing an upper bound and a lower bound for $\mathbb{E}[\tilde{r}_p]$. We first proceed with an upper bound in Lemma C.1.

**Lemma C.1.** *Let $X \subseteq \mathbb{R}^m$ and let $p \in X$. Let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X \cdot \log(1/\epsilon)/\epsilon^2)$. Let $r_p$ and $\widetilde{r}_p$ be the radius of $p$ and $Gp$ in $\mathbb{R}^m$ and $\mathbb{R}^d$ respectively, computed according to Eq. (3) in the main paper. Then*

$$\mathbb{E}[\tilde{r}_p] \leq (2 + O(\epsilon)) r_p.$$

*Proof.* Let $\delta > 0$ be fixed and let $\mathcal{E}_k$ be the event that

$$\max_{x \in B(p, r_p) \cap X} \|G(x - p)\| \in [(k-1)(1+\delta)r_p, k(1+\delta)r_p)$$

Note that $\mathcal{E}_k$ implies that there exists an $x \in B(p, r_p) \cap X$ such that $\|G(x-p)\| \geq (k-1)(1+\delta)r_p$, so by Lemma A.3 we have

$$\begin{aligned} \Pr(\mathcal{E}_k) &\leq \Pr(\exists x \in B(p, r_p) \cap X, \|G(x-p)\| \geq k(1+\delta)r_p) \\ &\leq \exp(-c(k-1)^2(1+\delta)^2 d) \end{aligned} \tag{5}$$

for some constant $c$. We now show that conditioned on $\mathcal{E}_k$, we have $\tilde{r}_p \leq (k+1)r_p(1+\delta)$. This is because conditioning on $\mathcal{E}_k$ gives us

$$\begin{aligned} &\sum_{Gq \in B(Gp, (k+1)(1+\delta)r_p)} ((k+1)r_p(1+\delta) - \|Gp - Gq\|) \\ &\qquad \geq \sum_{q \in B(p, r_p)} (1+\delta) r_p, \end{aligned}$$

where $Gq \in B(Gp, (k+1)(1+\delta)r_p)$ is interpreted as summing over the points in the set

$$GX \cap B(Gp, (k+1)(1+\delta)r_p).$$

Furthermore,

$$\sum_{q \in B(p,r_p)} r_p(1+\delta) \geq \sum_{q \in B(p,r_p)} r_p \geq \sum_{q \in B(p,r_p)} (r_p - \|p-q\|) = 1.$$

Therefore, by the observation that the function $f(r) = \sum_{q \in B(p,r)}(r - \|p-q\|)$ is increasing in $r$, it follows that $(k+1)(1+\delta)r_p \geq \tilde{r}_p$. Therefore, we have

$$\mathbb{E}[\tilde{r}_p \mid \mathcal{E}_k] \leq (k+1)(1+\delta)r_p. \tag{6}$$

Now using Eq. (5)

$$\begin{aligned}\mathbb{E}[\tilde{r}_p] &= \sum_{k=1}^{\infty} \mathbb{E}[\tilde{r}_p \mid \mathcal{E}_k] \Pr(\mathcal{E}_k) \\ &\leq (2+\delta)r_p \\ &\qquad + r_p \sum_{k=2}^{\infty}(k+1)\exp(-c(k-1)^2(1+\delta)^2 d) \\ &\leq (2+\delta)r_p \\ &\qquad + (1+\delta)r_p \int_0^{\infty} (x+2)\exp(-C'x^2)\, dx\end{aligned}$$

where $C' = c(1+\delta)^2 d$. We can explicitly evaluate that

$$\int_0^{\infty} (x+2)\exp(-C'x^2)\, dx = \sqrt{\frac{\pi}{C'}} + \frac{1}{2C'}.$$

Noting that $d = \Omega(1/\epsilon^2)$, we have that

$$\mathbb{E}[\tilde{r}_p] \leq (2 + O(\epsilon))r_p$$

by picking $\delta = O(\epsilon)$. $\square$

We now show the corresponding lower bound.

**Lemma C.2.** *Let $X \subseteq \mathbb{R}^m$ and let $p \in X$. Let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X \cdot \log(1/\epsilon)/\epsilon^2)$. Let $r_p$ and $\widetilde{r}_p$ be the radius of $p$ and $Gp$ in $\mathbb{R}^m$ and $\mathbb{R}^d$ respectively, computed according to Eq. (3) from the main paper. Then*

$$\mathbb{E}[\tilde{r}_p] \geq \frac{(1-\epsilon)r_p}{4}.$$

*Proof.* Let $k$ be the size of the set $|B(p, r_p/2) \cap X|$. By definition of $r_p$, the following inequality holds:

$$\begin{aligned}1 &= \sum_{q \in B(p,r_p)} (r_p - \|p-q\|) \\ &\geq \sum_{q \in B(p,r_p/2)} (r_p - \|p-q\|) \geq \frac{kr_p}{2}.\end{aligned} \tag{7}$$

Now let $\mathcal{E}$ be the event that the ball $B(Gp, (1-\epsilon)r_p/2)$ contains at most $k$ points. By invoking Theorem A.4, we will show that $\Pr(\mathcal{E}) \geq 1/2$. Consider the set $X$ without the $k-1$ points in $B(p, r_p/2) - \{p\}$, and with an extra point $q$ at distance $(1-\epsilon)r_p/2$ from $p$. The added point $q$ becomes a nearest neighbor of $p \in X$. By Theorem A.4 part (2) applied to an appropriately chosen $\epsilon' = O(\epsilon)$, with probability at least $1/2$, no point outside of $B(p, r_p/2)$ is mapped within $(1-\epsilon)r_p/2$ of $p$. After removing $q$, only the originally removed $k-1$ points (and $p$) can lie in $B(p, r_p(1-\epsilon)/2)$.

Conditioning on $\mathcal{E}$, it follows that

$$\sum_{Gq \in B(Gp,(1-\epsilon)r_p/2)} \left(\frac{(1-\epsilon)r_p}{2} - \|Gp - Gq\|\right) \leq \frac{(1-\epsilon)kr_p}{2}. \tag{8}$$

Combining Eq. (8) with (7), we have that

$$\sum_{Gq \in B(Gp,(1-\epsilon)r_p/2)} \left(\frac{(1-\epsilon)r_p}{2} - \|p-q\|\right) \leq 1 - \epsilon < 1.$$

Therefore, conditional on $\mathcal{E}$, it follows that $\widetilde{r}_p \geq (1-\epsilon)r_p/2$. Hence,

$$\mathbb{E}[\tilde{r}_p] \geq \frac{\mathbb{E}[\tilde{r}_p \mid \mathcal{E}]}{2} \geq \frac{(1-\epsilon)r_p}{4}. \qquad \square$$

Combining Lemma C.1 and C.2 gives us the complete proof of Theorem 4.1 in the main paper.

*Proof of Theorem 4.1 in the main paper.* The theorem follows from combining the result given in Lemma 3.1 from the main paper, that the sum of the radii $r_p$ is a constant factor approximation to the global optimal solution, and Lemmas C.1 and C.2 that state that $\mathbb{E}[\tilde{r}_p]$ is a constant factor approximation to $r_p$. $\square$

### C.2. Obtaining a Solution to Facility Location in Larger Dimension

Recall that the main technical challenge is to show that if a facility is within distance $O(\tilde{r}_p)$ of a fixed point $p$ in $\mathbb{R}^d$ (note that $\tilde{r}_p$ is calculated according to Eq. (3) from the main paper, in $\mathbb{R}^d$), then the facility must also be within distance $O(r_p)$ in $\mathbb{R}^m$, the larger dimension. We prove this claim formally in Theorem C.4.

Before presenting Theorem C.4, we need the following technical result later on for our probability calculations.

**Lemma C.3.** *Denote* $\mathrm{erf}(x)$ *to be the error function defined as*

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt.$$

*Then,*

$$1 - \mathrm{erf}(x) \leq \exp(-x^2)$$

*for all* $x \geq 1$.

*Proof.* Note that $f(x)\exp(-t^2)/\sqrt{\pi}$ is a valid probability density function over $\mathbb{R}$ so that

$$\operatorname{erf}(x) = 1 - \Pr(|Z| \geq x)$$

where $Z$ is distributed according to the density $f$. Now

$$\begin{aligned}\Pr(Z \geq x) &= \frac{1}{\sqrt{\pi}} \int_x^\infty \exp(-t^2)\, dt \\ &\leq \frac{1}{\sqrt{\pi}} \int_x^\infty \frac{t}{x} \exp(-t^2)\, dt \\ &= \frac{\exp(-x^2)}{2x\sqrt{\pi}}\end{aligned}$$

where the inequality follows from the fact that $t \geq x$. By symmetry, we have

$$\operatorname{erf}(x) + \exp(-x^2) - 1 \geq \exp(-x^2)\left(1 - \frac{1}{x\sqrt{\pi}}\right) \geq 0$$

for $x \geq 1$. □

The proof of Theorem C.4 relies on the careful balancing of the following two events. First, we control the value of the radius $\tilde{r}_p$ and show that $\tilde{r}_p \approx r_p$. In particular, we show that the probability of $\tilde{r}_p \geq kr_p$ for any constant $k$ is exponentially decreasing in $k$. The argument for this part follows similarly to the argument in Lemma C.1.

Next, we need to bound the probability that a 'far' point comes 'close' to $p$ after the dimensionality reduction. While Theorem A.4 roughly states that 'far' points do not come too 'close', we need a more detailed result to quantify how close far points can come after the dimension reduction.

To study this in a more refined manner, we bucket the points in $X \setminus \{p\}$ according to their distance from $p$. The distance spacing between buckets will be a linear scale. We show that points in $X \setminus \{p\}$ that are in 'level' $i$ do not shrink to a 'level' smaller than $O(\sqrt{i})$. Note that we need to control this even across all levels. To do this requires a chaining type argument which crucially depends on the doubling dimension of $X$. Finally, a careful combination of probabilities gives us our result.

**Theorem C.4.** *Let $X \subseteq \mathbb{R}^m$ and let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X \cdot \log(1/\epsilon)/\epsilon^2)$. Fix $p \in X$ and let $x \in X$ be the point that maximizes $\|p - x\|$ subject to the condition $Gx \in B(Gp, C\tilde{r}_p)$ where $C$ is a fixed constant. Then*

$$\mathbb{E}\|p - x\| \leq 2C(1 + O(\epsilon))r_p.$$

*Proof.* For simplicity, let $r = r_p, \tilde{r} = \tilde{r}_p$, and define $t_{-1} = 0, t_0 = 1$, and

$$t_i = 1 + 2\epsilon + \frac{\epsilon(i-1)}{4}$$

for all $i \geq 1$. Define $\mathcal{E}_i$ to be the event that $2Crt_i \leq \|p - x\| \leq 2Crt_{i+1}$ (the range $2Crt_i$ to $2Crt_{i+1}$ are our 'buckets' from the discussion preceding the proof). Then

$$\mathbb{E}\|p - x\| = \sum_{i \geq -1} \mathbb{E}[\|p - x\| \mid \mathcal{E}_i] \Pr(\mathcal{E}_i). \tag{9}$$

We first bound $\Pr(\mathcal{E}_i)$ in two different ways. By conditioning on the value of $\tilde{r}$, we can write this probability as

$$\begin{aligned}\Pr(\mathcal{E}_i) &= \sum_{j \geq -1} \Pr(\mathcal{E}_i \text{ and } 2rt_j \leq \tilde{r} \leq 2rt_{j+1}) &(10)\\ &= \sum_{j \geq -1} \big[\Pr(\mathcal{E}_i \mid 2rt_j \leq \tilde{r} \leq 2rt_{j+1}) \\ &\qquad \cdot \Pr(2rt_j \leq \tilde{r} \leq 2rt_{j+1})\big]. &(11)\end{aligned}$$

In the first of our two bounds for $\Pr(\mathcal{E}_i)$, we proceed by bounding $\Pr(2rt_j \leq \tilde{r} \leq 2rt_{j+1})$. Heuristically, the event $2rt_j \leq \tilde{r} \leq 2rt_{j+1}$ would mean that some point in $B(p, r)$ will be very far away from $p$ after the random projection and the probability of this event can be controlled very well.

More formally, we first claim that the event $2rt_j \leq \tilde{r} \leq 2rt_{j+1}$ implies that there exists a point $z$ in $B(p, r)$ such that $\|G(z - p)\| \geq rt_j$. This is because otherwise, we have $\|Gp - Gq\| < rt_j$ for all $q \in B(p, r)$. This means that

$$\begin{aligned}\sum_{q \in B(Gp, 2rt_j)} (2rt_j - \|Gp - Gq\|) &> \sum_{q \in B(p, rt_j)} (2rt_j - rt_j) \\ &\geq |B(p, r) \cap X| \cdot rt_j \\ &\geq |B(p, r) \cap X| \cdot r.\end{aligned}$$

We also know that $|B(p, r) \cap X| \cdot r \geq 1$ from (4). Altogether, we have that $\sum_{q \in B(Gp, 2rt_j)} (2rt_j - \|Gp - Gq\|) > 1$ which cannot happen by definition of $\tilde{r}$ and our assumption that $2rt_j \leq \tilde{r}$ (see Figure 2 in the main paper). Therefore by Lemma A.3, we have

$$\Pr(2rt_j \leq \tilde{r} \leq 2rt_{j+1}) \leq \exp(-C_1 d t_j^2) \tag{12}$$

for some constant $C_1$. Summing over the variable $j$ in inequality (12) gives us a bound on $\Pr(\mathcal{E}_i)$. We will only end up using this bound for $j \geq \Omega(\sqrt{i})$, and will use the second bound for small $j$.

We now give a second bound on $\Pr(\mathcal{E}_i)$ by controlling $\Pr(\mathcal{E}_i \text{ and } 2rt_j \leq \tilde{r} \leq 2rt_{j+1})$. Note that the event $\mathcal{E}_i$ and $2rt_j \leq \tilde{r} \leq 2rt_{j+1}$ together imply that there exists some $x$ that satisfies

$$2Crt_i \leq \|p - x\| \leq 2Crt_{i+1} \text{ and } \|G(x - p)\| \leq 2Crt_{j+1}$$

due to the fact that $Gx$ is a point in $B(Gp, C\tilde{r}_p)$. Therefore,

$$\begin{aligned}&\Pr(\mathcal{E}_i \text{ and } 2rt_j \leq \tilde{r} \leq 2rt_{j+1}) \\ &\leq \Pr(\exists x,\, 2Crt_i \leq \|p - x\| \leq 2Crt_{i+1} \\ &\qquad \text{and } \|G(x - p)\| \leq 2Crt_{j+1}).\end{aligned}$$

We bound the right hand side of the above probability for the range $j = O(\sqrt{i})$. Let

$$X_i = \{x \in X \mid 2Crt_i \le \|x - p\| < 2Crt_{i+1}\}.$$

By the definition of doubling dimension, we can find a covering of $X_i$ with at most $\lambda^{O(\log(t_i/\epsilon))}$ balls of radius $2Cr\epsilon/4$ centered at points in some set $S \subseteq X$. Then by Lemma A.3, we have

$$\Pr\Bigg(\exists s \in S \; \exists x \in B(s, 2Cr\epsilon/4) \cap X_i, \\ \|Gs - Gx\| \ge \frac{2Cr\epsilon\sqrt{i}}{8}\Bigg) \le \exp(-O(di)) \quad (13)$$

if $d \ge \Omega(\log(\lambda)\log(1/\epsilon)/\epsilon^2)$. Now fix $s \in S$. If $\|G(s - p)\| < 2Cr(1 + \epsilon + \epsilon\sqrt{i}/4)$ then

$$\frac{\|G(s-p)\|}{\|s-p\|} \le \frac{1+\epsilon+\epsilon\sqrt{i}/4}{1+2\epsilon+\epsilon i/4} \le \begin{cases} 1-\epsilon/4, & \text{for } 0 \le i \le 1/\epsilon^2 \\ O(1)/\sqrt{i}, & \text{for } i > 1/\epsilon^2 \end{cases}.$$

Hence by applying the two inequalities (1) and (2) to the unit vector $(s - p)/\|s - p\|$, we have

$$\Pr\left(\exists s \in S, \|G(s-p)\| \le 2Cr\left(1+\epsilon+\frac{\epsilon\sqrt{i}}{4}\right)\right) \le \begin{cases} \exp(-c^{''} d\epsilon^2), & \text{for } 0 \le i \le 1/\epsilon^2 \\ i^{-c^{''} d}, & \text{for } i > 1/\epsilon^2 \end{cases}.$$

Note that we used the inequality (1) for the bound $i \le 1/\epsilon^2$ and the inequality (2) for $i > 1/\epsilon^2$. Combining the above bound with the inequality in (13) gives us

$$\Pr\left(\exists x \in X_i, \|G(x-p)\| \le 2Cr\left(1+\epsilon+\frac{\epsilon\sqrt{i}}{8}\right)\right) \le \begin{cases} 2\exp(-c^{''} d\epsilon^2), & \text{for } 0 \le i \le 1/\epsilon^2 \\ 2i^{-c^{''} d}, & \text{for } i > 1/\epsilon^2 \end{cases}.$$

Thus for $j \le C_2\sqrt{i}$, we have

$$\Pr(\exists x, 2Crt_{j+1} > \|G(x-p)\| \text{ and } 2Crt_i \le \|x-p\| < 2Crt_{i+1}) \le \begin{cases} 2\exp(-C_3 d\epsilon^2), & \text{for } 0 \le i \le 1/\epsilon^2 \\ 2i^{-C_3 d}, & \text{for } i > 1/\epsilon^2 \end{cases}$$

where $C_2, C_3$ are fixed constants. Using the representation given in (11) for $\Pr(\mathcal{E}_i)$ along with (12), we see that for $0 \le i \le 1/\epsilon^2$, we can bound

$$\Pr(\mathcal{E}_i) \le 4\exp(-C_2 d\epsilon^2) + \sum_{j\ge1}\exp(-C_1 dj^2\epsilon^2) \quad (14)$$

while for $i > 1/\epsilon^2$, we instead use the following stronger bound

$$\Pr(\mathcal{E}_i) \le 2C_2\sqrt{i}\cdot i^{-C_3 d} + \sum_{j\ge C_2\sqrt{i}}\exp(-C_1 dj^2\epsilon^2) \quad (15)$$

which comes from using (10) for $j \le C_2\sqrt{i}$ and (11) for larger $j$. Combing these bounds with (9), we have

$$\mathbb{E}\|p - x\| \le 2C(1+O(\epsilon))r + \sum_{i\ge0} 2Crt_{i+1}\Pr(\mathcal{E}_i). \quad (16)$$

Our task is to now bound the sum $\sum_{i\ge0} t_{i+1}\Pr(\mathcal{E}_i)$. In the rest of the proof, we will show that this sum is $O(\epsilon)$. We split the sum into two terms depending on if $i \le 1/\epsilon^2$ or if $i > 1/\epsilon^2$. Using the bounds (14) and (15) gives us

$$\sum_{i\ge0} t_{i+1}\Pr(\mathcal{E}_i) \le C_4\sum_{i\le1/\epsilon^2} i\left(\exp(-C_2 d\epsilon^2) + \sum_{j\ge1}\exp(-C_1 dj^2\epsilon^2)\right) \quad (17)$$

$$+ C_4\sum_{i>1/\epsilon^2}\epsilon i\left(i^{-C_2 d+1/2} + \sum_{j\ge C_2\sqrt{i}}\exp(-C_1 dj^2\epsilon^2)\right) \quad (18)$$

for some constant $C_4$. In (17), we are using the fact that $t_{i+1} = O(i)$ and for (18), we are instead using $t_{i+1} = O(\epsilon i)$. We can bound (17)

$$\frac{\exp(-C_2 d\epsilon^2)}{\epsilon^4} + \frac{1}{\epsilon^2}\sum_{j\ge1}\exp(-C_1 dj^2\epsilon^2) \le O(\epsilon) \quad (19)$$

by using the fact that $d = \Omega(\log(1/\epsilon)/\epsilon^2)$.

We now focus on bounding (18). As a first step, we have the estimate

$$\epsilon\sum_{i>1/\epsilon^2} i^{-C_2 d+3/2} = O(\epsilon)$$

which holds for large enough constant $d$. Finally, bounding the remaining sum of (18) by an integral gives us

$$\epsilon\sum_{i>1/\epsilon^2} i \sum_{j\ge C_2\sqrt{i}}\exp(-C_1 dj^2\epsilon^2) \le \epsilon\int_1^\infty x\int_{\sqrt{x}}^\infty \exp(-C_5 dt^2\epsilon^2)\,dt\,dx$$

for some constant $C_5$. Now using the definition of the complementary error function, we can compute that

$$\epsilon\int_1^\infty x\int_{\sqrt{x}}^\infty \exp(-C_5 dt^2\epsilon^2)\,dt\,dx \le O(d^{-1/2})\int_1^\infty x\cdot\text{erfc}(\epsilon\sqrt{C_5 d\cdot x})\,dx \le O(\epsilon)\int_1^\infty x\cdot\text{erfc}(\epsilon\sqrt{C_5 d\cdot x})\,dx.$$

From Lemma C.3, we have

$$\int_1^\infty x \cdot \text{erfc}(\epsilon\sqrt{C_5 d \cdot x})\, dx \le \int_1^\infty x \cdot \exp(-C_5^2 \epsilon^2 d \cdot x)\, dx = O(\epsilon) \quad (20)$$

using the fact that $d = \Omega(\log(1/\epsilon)/\epsilon^2)$. Altogether, the bounds (19) and (20) allow us to bound the right hand side of (17) and (18) and therefore, bound the sum $\sum_{i \ge 0} t_{i+1} \Pr(\mathcal{E}_i)$ as $O(\epsilon)$. Finally, using (16), we end up with

$$\mathbb{E}\|p - x\| \le 2C(1 + O(\epsilon))r. \qquad \square$$

As a corollary, we can prove Theorem 4.2 in the main paper.

*Proof of Theorem 4.2 in the main paper.* Let $\mathcal{F}_d$ be a locally optimal solution in $GX$. When we evaluate the cost of $\mathcal{F}_d$ in the larger dimension $\mathbb{R}^m$, the number of facilities stays the same. Now since $\mathcal{F}_d$ is a locally optimal solution in $\mathbb{R}^d$, each point $p$ has a facility that is within distance $C\tilde{r}_p$ in $\mathbb{R}^d$. Then by Theorem C.4, the connection cost of $p$ in the larger dimension is bounded by $C' r_p$, for some constant $C'$, in expected value. Summing over all points $p \in X$ gives us

$$\mathbb{E}[\text{cost}_m(\mathcal{F}_d)] \le |\mathcal{F}_d| + O\left(\sum_{p \in X} r_p\right).$$

Finally, since $|\mathcal{F}_d| \le \text{cost}_d(\mathcal{F}_d)$ by definition, and since $\sum_{p \in X} r_p = O(F)$ by Lemma 3.1 in the main paper, we have that

$$|\mathcal{F}_d| + O\left(\sum_{p \in X} r_p\right) \le \text{cost}_d(\mathcal{F}_d) + O(F).$$

Together, these prove the main theorem of this section. $\square$

## D. Dimension Reduction for MST: Omitted Proofs

In this section, we prove Lemma 5.2 and Theorem 5.1 from the main paper.

### D.1. Proof of Theorem 5.1 (from the main paper)

In this subsection, we prove that Lemma 5.2 (from the main paper) implies Theorem 5.1 (from the main paper). To see why, first note that $\text{cost}_X(\widetilde{\mathcal{M}}) \ge \text{cost}_X(\mathcal{M})$ and $\text{cost}_{GX}(\mathcal{M}) \ge \text{cost}_{GX}(\widetilde{\mathcal{M}})$, since $\mathcal{M}$ is the minimum spanning tree on $X$ and $\widetilde{\mathcal{M}}$ is the minimum spanning tree on $GX$. Moreover, for each edge $e = (x, y) \in \mathcal{M}$, $\|Gx - Gy\|$ has distribution $\chi_d/\sqrt{d} \cdot \|x - y\|$, where $\chi_d$ is the square root of a chi-square with $d$ degrees of freedom. This has mean

$$\mu = \|x-y\| \cdot \frac{1}{\sqrt{2d}} \cdot \frac{\Gamma((d+1)/2)}{\Gamma(d/2)} = \|x-y\| \cdot \left(1 - O\left(\frac{1}{d}\right)\right)$$

and variance $\|x - y\|^2 - \mu^2 = \|x - y\|^2 \cdot O(1/d)$ (Wolfram Research). Therefore, the standard deviation of $\|G(x - y)\|$ is at most $\epsilon \cdot \|x - y\|$ since $d = \Omega(\epsilon^{-2})$. Therefore, the expectation of $\text{cost}_{GX}(\mathcal{M})$ is $\sum_{e=(x,y) \in \mathcal{M}} \|x - y\| \cdot (1 - O(1/d)) = M \cdot (1 - O(1/d))$. Also, using the well known fact that for any (possibly correlated) random variables $X_1, \ldots, X_n$, $\sqrt{Var(X_1 + \cdots + X_n)} \le \sum \sqrt{Var(X_i)}$, we have that the standard deviation of $\text{cost}_{GX}(\mathcal{M})$ is at most $\sum_{e=(x,y) \in \mathcal{M}} \epsilon \cdot \|x - y\| = \epsilon \cdot M$.

To finish, define random variables $Z_1 = \text{cost}_X(\widetilde{\mathcal{M}}) - \text{cost}_X(\mathcal{M})$, $Z_2 = \text{cost}_X(\mathcal{M}) - \text{cost}_{GX}(\mathcal{M})$, and $Z_3 = \text{cost}_{GX}(\mathcal{M}) - \text{cost}_{GX}(\widetilde{\mathcal{M}})$. Our observations from the previous paragraph tell us that $Z_1$ and $Z_3$ are nonnegative, and $Z_2$ has nonnegative expectation and standard deviation bounded by $O(\epsilon) \cdot M$. Finally, Lemma 5.2 (from the main paper) tells us that $\mathbb{E}[Z_1 + Z_2 + Z_3] \le O(\epsilon) \cdot M$. However, this means that $\mathbb{E}[Z_1] \le O(\epsilon) \cdot M$, so $0 \le Z_1 \le O(\epsilon) \cdot M$ with high probability by Markov's inequality. Therefore, $\text{cost}_X(\widetilde{\mathcal{M}}) \le (1 + O(\epsilon)) \cdot M$ with high probability, so the pullback is a $1 + O(\epsilon)$ approximation with high probability. Likewise, we also have that $0 \le Z_3 = O(\epsilon)$ with high probability, and since $\mathbb{E}[Z_2]$, $\sqrt{Var(Z_2)} \le O(\epsilon) \cdot M$, we also have that $|Z_2| = O(\epsilon)$ with high probability. Thus, $|Z_2 + Z_3| = O(\epsilon)$ with high probability, which means $\text{cost}_{GX}(\widetilde{\mathcal{M}}) \in [1 - O(\epsilon), 1 + O(\epsilon)] \cdot M$. As a result, the MST cost is preserved under dimensionality reduction with high probability as well.

### D.2. Proof of Lemma 5.2 (from the main paper)

In this subsection, we prove prove Lemma 5.2 (from the main paper). In fact, we show the following stronger statement.

$$\mathbb{E}_G\left[\sum_{e=(x,y) \in \widetilde{\mathcal{M}}} \max(0, \|x - y\| - (1 + 5\epsilon)\|Gx - Gy\|)\right] \le \epsilon \cdot M. \quad (21)$$

To see why this implies Lemma 5.2 (from the main paper), by removing the maximum with 0, Equation (21) implies that $\mathbb{E}_G[\text{cost}_X(\widetilde{\mathcal{M}}) - (1 + 5\epsilon) \cdot \text{cost}_{GX}(\widetilde{\mathcal{M}})] \le \epsilon \cdot M$. But $\mathbb{E}_G[\text{cost}_{GX}(\widetilde{\mathcal{M}})] \le \mathbb{E}_G[\text{cost}_{GX}(\mathcal{M})] \le (1 + \epsilon) \cdot M$, which means that $\mathbb{E}_G[\text{cost}_X(\widetilde{\mathcal{M}}) - \text{cost}_{GX}(\widetilde{\mathcal{M}})] \le \epsilon \cdot M + 5\epsilon \cdot (1 + \epsilon) \cdot M = O(\epsilon) \cdot M$.

*Proof of Equation* (21). Consider some range $A_i = \left[(1 + \epsilon)^i, (1 + \epsilon)^{i+1}\right)$. We will bound the expectation of

$$K_i := \sum_{\substack{e=(x,y) \in \widetilde{\mathcal{M}} \\ \|Gx - Gy\| \in A_i}} \max\left(0, \|x - y\| - (1 + 5\epsilon)\|Gx - Gy\|\right)$$

and sum our upper bounds for $\mathbb{E}_G[K_i]$ over a range of $i$. For $K_i$ to be nonzero, we need there to exist $(x, y)$ such that $\|x - y\| \geq \|Gx - Gy\|$ and $\|Gx - Gy\| \in A_i$, so $\|x - y\| \geq (1 + \epsilon)^i$. Therefore, we only need to sum $\mathbb{E}[K_i]$ over integers $i$ such that $(1 + \epsilon)^i \leq \operatorname{diam}(X)$.

To do this, first consider some fixed $i$ and some sufficiently large constant $C_1$, and define $t := t_i := \frac{\epsilon}{C_1} \cdot (1 + \epsilon)^i$. Consider the following greedy procedure of selecting a partition of $X$. First, pick some point $x_1$ arbitrarily, then pick some point $x_2$ of distance more than $t$ from $x_1$ (in the original space), then some point $x_3$ of distance more than $t$ from $x_1$ and $x_2$, and so on until we have some $x_1, \dots, x_r$ and can no longer pick any more points. Finally, we partition $X$ into subsets $X_1, \dots, X_r$ so that each $x \in X$ is in $X_p$ if $x_p$ is the closest point to $x$ (breaking ties arbitrarily). Note that the partitioning is deterministic (independent of $G$). We show the following proposition:

**Proposition D.1.** *The MST cost $M$ of the dataset $X$ (in the original space $\mathbb{R}^m$) is at least $\frac{r \cdot t}{2}$.*

*Proof.* By a known result on Steiner Trees (Kou et al., 1981), $M$ is at least $\frac{r}{2(r-1)}$ times the MST cost of the set $\{x_1, \dots, x_r\} \subset X$, assuming $r \geq 2$. As the distance between any $x_p, x_q$ is at least $r$, the MST cost of $\{x_1, \dots, x_r\}$ is at least $(r - 1) \cdot t$, so $M \geq \frac{r}{2(r-1)} \cdot (r - 1) \cdot t = \frac{r \cdot t}{2}$. Finally, as $(1+\epsilon)^i \leq \operatorname{diam}(X)$, we have $t \leq \operatorname{diam}(X)/C_1$, so if $C_1 > 2$, then the greedy procedure of partitioning $X$ cannot end with just $x_1$, so indeed $r \geq 2$. ☐

Now, we consider partitioning each $X_p$ into subsets $X_{p,1}, \dots, X_{p,s}$ as follows. Since the radius of $X_p$ is at most $t$, by definition of the doubling dimension, for each $k \geq 1$ we can split $X_p$ into at most $\lambda_X^k$ balls of radius at most $t/2^k$. We choose the smallest integer $k$ so that all of these balls have diameter at most $2t$ when projected by $G$, and let $s_p = s$ be the number of subsets $X_{p,q}$ formed for each $p$. (Note: this partitioning $X_{p,q}$ is now dependent on $G$.) We claim the following:

**Proposition D.2.** *For any fixed $p$ and all integers $k \geq 1$, $\Pr(s_p > \lambda_X^k) \leq \exp\left(-cd2^k\right)$.*

*Proof.* For any fixed $k$, we split $X_p$ into at most $\lambda_X^k$ balls of radius at most $t/2^k$: this process is independent of $G$. Now, fix a small ball: when we apply the random projection $G$, the probability that it has radius more than $t$ when projected is at most $\exp\left(-cd2^{2k}\right)$, by Lemma A.3. But there are $\lambda_X^k \leq \exp\left(cdk\right)$ such balls if $d$ is at least $c^{-1}\log\lambda_X$, so the probability that even one of $GX_{p,q}$ has radius more than $t$ is at most $\exp\left(-cd2^{2k}\right) \cdot \exp\left(cdk\right) = \exp\left(-cd(2^{2k} - k)\right) \leq \exp\left(-cd2^k\right)$. ☐

We also make the following observations:

1. If $x \in X_p$, then $\|x - x_p\| \leq t$, so the diameter of each $X_p$ is at most $2t$. Likewise, the diameter of each $X_{p,q}$ is at most $2t$ in both the original space and the reduced space.

2. By properties of the doubling dimension, for any $x_p$ and all $k \geq 1$, there are at most $\lambda_X^{C_2 \cdot k}$ points $\{x_{p'}\}_{p'=1}^r$ within $2^k \cdot t$ of $x_p$ for some $C_2$, since $x_1, \dots, x_r$ are all at least $t$ apart.

Recall that $D(X_p, X_{p'})$ is the *maximum* distance between points in $X_p$ and $X_{p'}$ (in the original space), as opposed to $d(X_p, X_{p'})$ which is the *minimum* distance. Now, for any fixed $i$, we bound the expectation of

$$L_i := \sum_{\substack{p,p' \\ d(GX_p, GX_{p'}) < (1+\epsilon)^{i+1} \\ D(X_p, X_{p'}) \geq (1+5\epsilon)\cdot(1+\epsilon)^i}} D(X_p, X_{p'}) \cdot s_p s_{p'},$$

where the sum is over all pairs $p, p' \in [r]$.

First, we make the following claim.

**Lemma D.3.** *For all $i$ and any fixed $G$, $L_i \geq K_i$.*

*Proof.* For any edge $e \in \widetilde{\mathcal{M}}$, if $e$ has length in range $A_i$ (in the projected space), then this length is greater than $2C_1 \cdot t$ (assuming $\epsilon < 1/2$). Then, $e$ is some edge $(Gx, Gy)$ where $x \in X_{p,q}, y \in X_{p',q'}$, where $(p, q) \neq (p', q')$ by Observation 1. So, if edge $e$ contributes toward the sum in $K_i$, then $D(X_p, X_{p'}) \geq \|x-y\| \geq (1+5\epsilon)\cdot\|Gx-Gy\| \geq (1 + 5\epsilon) \cdot (1 + \epsilon)^i$. At the same time, $d(GX_p, GX_{p'}) \leq \|Gx-Gy\| < (1+\epsilon)^{i+1}$. Thus, this pair $(p, p')$ contributes toward the sum in $L_i$. Moreover, $D(X_p, X_{p'}) \geq \|x - y\| \geq \max(0, \|x-y\| - (1+5\epsilon)\|Gx-Gy\|)$. This will be useful since $L_i$ is a sum over $D(X_p, X_{p'})$ (multiplied by $s_p s_{p'}$) and $K_i$ is a sum over $\max(0, \|x-y\| - (1+5\epsilon)\|Gx-Gy\|)$.

Finally, it is impossible for two pairs $(Gx, Gy)$ and $(Gx', Gy')$ to both be edges in $\widetilde{\mathcal{M}}$ that contribute to the sum $K_i$, if $x, x' \in X_{p,q}$ and $y, y' \in X_{p',q'}$. If there were such pairs $(Gx, Gy), (Gx', Gy')$, this means that the edges $(Gx, Gy)$ and $(Gx', Gy')$ have length in $A_i$, and therefore have length at least $2C_1 \cdot t$. However, the diameters of $GX_{p,q}$ and $GX_{p',q'}$ are at most $2t$, so it would be better to replace edge $(Gx', Gy')$ with either edge $(Gx, Gx')$ or edge $(Gy, Gy')$: exactly one of these replacements will preserve the spanning tree property, and either replacement reduces the total cost. Thus, for each pair $(p, p')$ contributing to the sum in $L_i$, at most $s_p \cdot s_{p'}$ corresponding pairs $(x, y)$ can contribute to the sum in $K_i$, and since $D(X_p, X_{p'}) \geq \max(0, \|x - y\| - (1 + 5\epsilon)\|Gx - Gy\|)$ whenever $x \in X_{p,q}, y \in X_{p',q'}$, this finishes the proof. ☐

We will now bound the expectation of $L_i$.

**Lemma D.4.** *For any fixed $i$, $\mathbb{E}[L_i] \leq \frac{\epsilon^2}{10 \log n} \cdot M$.*

*Proof.* For each $j \geq 1$, define $B_{i,j}$ to be the interval $[(1+5\epsilon)\cdot(1+\epsilon)^{i+j-1}, (1+5\epsilon)\cdot(1+\epsilon)^{i+j})$. Fix some $p, p'$ such that $D(X_p, X_{p'}) \in B_{i,j}$ (note: this is independent of $G$). Since all points in $X_p$ are at most $t$ from $x_p$ (and similar for $X_{p'}$), we have that $\|x_p - x_{p'}\| \geq (1+5\epsilon)\cdot(1+\epsilon)^{i+j-1} - 2t \geq (1+3\epsilon)\cdot(1+\epsilon)^{i+j}$. Now, if $d(GX_p, GX_{p'}) < (1+\epsilon)^{i+1}$, then one of the following three events must be true:

1. $\|x_p - x_{p'}\| \leq (1+\epsilon)^{i+(j/2)} \cdot (1+3\epsilon)$
2. $\mathrm{diam}(GX_p) \geq \epsilon \cdot (1+\epsilon)^{i+(j/2)}$
3. $\mathrm{diam}(GX_{p'}) \geq \epsilon \cdot (1+\epsilon)^{i+(j/2)}$.

Indeed, if all three were false, then $d(GX_p, GX_{p'}) \geq \|x_p - x_{p'}\| - \mathrm{diam}(GX_p) - \mathrm{diam}(GX_{p'}) \geq (1+\epsilon)^{i+(j/2)}\cdot(1+\epsilon) \geq (1+\epsilon)^{i+1}$.

Now, the probability of the first event (over the randomness of $G$) is at most the probability that a random projection shrinks $x_p - x_{p'}$ by a factor of at least $(1+\epsilon)^{j/2}$. By Equation (1), if $j \leq \epsilon^{-1}$, then this happens with probability at most $\exp\left(-d(j\epsilon)^2/100\right)$, and by Equation (2), if $j > \epsilon^{-1}$, then this happens with probability at most $(1+\epsilon)^{-(j/2)\cdot d/20} \leq \exp\left(-d(j\epsilon)/100\right)$. The probability of each of the second and third events occurring, since $\mathrm{diam}(X_p), \mathrm{diam}(X_{p'}) \leq \epsilon \cdot (1+\epsilon)^i / C_1$, is at most $\exp\left(-cd \cdot C_1^2 (1+\epsilon)^j\right) \leq \exp\left(-d(j\epsilon)/100\right)$ by Lemma A.3. Next, note that by Proposition D.2, for some constant $C_3$, $\Pr(s_p \geq \lambda_X^{C_3 \cdot k}) \leq \exp(-d \cdot 2^k/100)$ for all real $k \geq 1$, and the same is true for $s_{p'}$.

Again consider some fixed $j$ and some $p, p'$ with $D(X_p, X_{p'}) \in B_{i,j}$. Define the random variable $Z_{p,p'} := s_p s_{p'} \cdot \mathbb{I}\left(d(GX_p, GX_{p'}) < (1+\epsilon)^{i+1}\right)$, where $\mathbb{I}$ represents an indicator random variable. Then, if $j \leq \epsilon^{-1}$, $d(GX_p, GX_{p'}) < (1+\epsilon)^{i+1}$ occurs with probability at most $3 \cdot \exp\left(-d(j\epsilon)^2/100\right) \leq 3 \cdot \exp\left(-d \cdot \epsilon^2/100\right)$, so $\Pr(Z_{p,p'} > 0) \leq 3 \cdot \exp\left(-d \cdot \epsilon^2/100\right)$. Next, for any $k \geq 1$, if $Z_{p,p'} \geq \lambda_X^{2k \cdot C_3}$, then either $s_p$ or $s_{p'}$ is at least $\lambda_X^{k \cdot C_3}$, which occurs with probability at most $2\exp\left(-d \cdot 2^k/100\right)$ by Proposition D.2. Hence,

$$\begin{aligned}\mathbb{E}[Z_{p,p'}] &\leq 3 \cdot \exp\left(-\frac{d\epsilon^2}{100}\right) \cdot \lambda_X^{2 \cdot C_3} \\ &\quad + \sum_{k=1}^{\infty} 2 \cdot \exp\left(-\frac{d \cdot 2^k}{100}\right) \cdot \lambda_X^{2(k+1)\cdot C_3} \\ &\leq 10 \cdot \exp\left(-\frac{d\epsilon^2}{200}\right)\end{aligned}$$

by our choice of the dimension $d$. However, if $j > \epsilon^{-1}$, then $d(GX_p, GX_{p'})$ occurs with probability at most $3 \cdot \exp\left(-d(j\epsilon)/100\right)$, so $\Pr(Z_{p,p'} > 0) \leq 3 \cdot \exp\left(-d \cdot (j\epsilon)/100\right)$. But for any $k \geq 1$, if $Z_{p,p'} \geq \lambda_X^{2(k+\log(j\epsilon))\cdot C_3}$, then either $s_p$ or $s_{p'}$ is at least $\lambda_X^{(k+\log(j\epsilon))\cdot C_3}$, which occurs with probability at most $2\exp\left(-d \cdot 2^k \cdot (j\epsilon)/100\right)$. Hence,

$$\begin{aligned}\mathbb{E}[Z_{p,p'}] &\leq 3 \cdot \exp\left(-\frac{d(j\epsilon)}{100}\right) \cdot \lambda_X^{2(1+\log(j\epsilon))\cdot C_3} \\ &\quad + \sum_{k=1}^{\infty} 2 \cdot \exp\left(-\frac{d \cdot (j\epsilon) \cdot 2^k}{100}\right) \cdot \lambda_X^{2(k+1+\log(j\epsilon))\cdot C_3} \\ &\leq 10 \cdot \exp\left(-\frac{d(j\epsilon)}{200}\right)\end{aligned}$$

by our choice of the dimension $d$.

Next, note that for each $p$, the number of $p'$ with $D(X_p, X_{p'}) \leq (1+5\epsilon)\cdot(1+\epsilon)^{i+j} \leq \left[C_1 \cdot \epsilon^{-1} \cdot (1+\epsilon)^{5+j}\right] \cdot t$ is at most $\lambda_X^{C_2 \cdot (\log C_1 + \log \epsilon^{-1} + \epsilon\cdot(5+j))}$ by Observation 2. Hence, the total number of pairs $(p, p')$ with $D(X_p, X_{p'}) \in B_{i,j}$ is at most $r \cdot \lambda_X^{C_2 \cdot (\log C_1 + \log \epsilon^{-1} + \epsilon \cdot (5+j))} \leq r \cdot \lambda_X^{C_4 \cdot (\log \epsilon^{-1} + j\epsilon)}$ for some constant $C_4$.

Combining everything together, we have that

$$\begin{aligned}\mathbb{E}[L_i] &= \sum_{j \geq 1} \sum_{p,p' : D(X_p, X_{p'}) \in B_{i,j}} D(X_p, X_{p'}) \cdot \mathbb{E}[Z_{p,p'}] \\ &\leq \sum_{j=1}^{\epsilon^{-1}} \Bigg( r \cdot \lambda_X^{C_4 \cdot (\log \epsilon^{-1} + j\epsilon)} \cdot (1+5\epsilon)\cdot(1+\epsilon)^{i+j} \\ &\qquad \cdot 10 \cdot \exp\left(-\frac{d\epsilon^2}{200}\right)\Bigg) \\ &\quad + \sum_{j > \epsilon^{-1}} \Bigg( r \cdot \lambda_X^{C_4 \cdot (\log \epsilon^{-1} + j\epsilon)} \cdot (1+5\epsilon)\cdot(1+\epsilon)^{i+j} \\ &\qquad \cdot 10 \cdot \exp\left(-\frac{d(j\epsilon)}{200}\right)\Bigg) \\ &\leq 20 C_1 r t \cdot \Bigg( \sum_{j=1}^{\epsilon^{-1}} \lambda_X^{C_5 (\log \epsilon^{-1})} \exp\left(-\frac{d\epsilon^2}{200}\right) \qquad (22) \\ &\qquad + \sum_{j > \epsilon^{-1}} \lambda_X^{C_5 (\log \epsilon^{-1} + j\epsilon)} \exp\left(-\frac{d(j\epsilon)}{200}\right)\Bigg) \qquad (23)\end{aligned}$$

for some constant $C_5$. Above, the first equality follows by definition of $L_i$. The next inequality follows from our bound on $\mathbb{E}[Z_{p,p'}]$, our bound the number of $(p, p')$ with $D(X_p, X_{p'}) \in B_{i,j}$, and since $D(X_p, X_{p'}) \in B_{i,j}$ implies $D(X_p, X_{p'}) \leq (1+5\epsilon)\cdot(1+\epsilon)^{i+j}$. The final inequality follows from simple factorization and the facts that $C_1 t \leq (1+\epsilon)^i$ and $1 + 5\epsilon \leq 2$.

Now, if we choose $d = C_6 \cdot (\log\log n + \log \epsilon^{-1} \log \lambda_X) \cdot \epsilon^{-2}$ for some sufficiently large constant $C_6$, we have that

$\lambda_X^{C_5(\log \epsilon^{-1})} \cdot \exp(-d\epsilon^2/200) \le \frac{\epsilon^3}{1000C_1 \log n}$ for all $j \le \epsilon^{-1}$, and $\lambda_X^{C_5(\log \epsilon^{-1}+j\epsilon)} \cdot \exp\left(-d(j\epsilon)/200\right) \le \exp\left(j \cdot \epsilon^{-1}\right) \cdot \exp\left(-d(j\epsilon)/400\right) \le \frac{\exp(-j)}{1000C_1 \log n}$ for all $j > \epsilon^{-1}$. Hence, Equation (23) can be upper bounded by

$$20C_1 \cdot rt \cdot \left(\epsilon^{-1} \cdot \frac{\epsilon^3}{1000C_1 \log n} + \frac{\sum_{j \ge \epsilon^{-1}} e^{-j}}{1000C_1 \log n}\right) \le \frac{\epsilon^2}{20 \log n} \cdot rt. \quad (24)$$

Proposition D.1 tells us that $rt \le 2M$. Hence, Equation (24) is at most $\frac{\epsilon^2}{10 \log n} \cdot M$, as desired. $\square$

In sum, we have that $\mathbb{E}[K_i] \le \frac{\epsilon^2}{10 \log n} \cdot M$ for all $i$. Moreover, for small $i$, Equation (24) tells us that $\mathbb{E}[K_i] \le \frac{\epsilon^2}{20 \log n} \cdot rt \le \frac{\epsilon^2}{20 \log n} \cdot n \cdot (1+\epsilon)^i$, since $r \le n$ and $t \le (1+\epsilon)^i$. Therefore, the LHS of Equation (21) is at most

$$\sum_{i:(1+\epsilon)^i \le \operatorname{diam}(X)} \mathbb{E}[K_i] \le \frac{\epsilon^2}{20 \log n} \cdot \sum_{i:(1+\epsilon)^i \le \operatorname{diam}(X)} \min\left(2M, n \cdot (1+\epsilon)^i\right).$$

Using the bound $\frac{\epsilon^2}{10 \log n} \cdot M$ for $i$ with $\frac{\operatorname{diam}(X)}{n} < (1+\epsilon)^i \le \operatorname{diam}(X)$ and the bound $\frac{\epsilon^2}{20 \log n} \cdot n \cdot (1+\epsilon)^i$ for $i$ with $(1+\epsilon)^i \le \frac{\operatorname{diam}(X)}{n}$, we can bound this by

$$\frac{\epsilon^2}{20 \log n} \cdot \left(2M \cdot \log_{1+\epsilon} n + \frac{\operatorname{diam}(X)}{1 - \frac{1}{1+\epsilon}}\right) \le \frac{\epsilon \cdot M}{5} + \frac{\operatorname{diam}(X) \cdot \epsilon}{10 \log n} < \epsilon \cdot M,$$

since $\operatorname{diam}(X) \le M$. This concludes the proof. $\square$

# E. Lower Bounds: Omitted Proofs

## E.1. Dependence on the Doubling Dimension

In this subsection, we prove Theorems 6.1, 6.2, and 6.3 from the main paper.

We begin with Theorem 6.1 (from the main paper). To do so, we construct a set $X$ of $m$ points in $\mathbb{R}^m$ such that if we randomly project $X$ to $o(\log m)$ dimensions, then with high probability, the facility location cost is not preserved up to a constant factor. Moreover, the optimal set of facility centers in the projected space, with high probability, is not a constant-factor approximation to facility location in the original space. The point set $X$ we choose will just be a scaled set of identity vectors in $\mathbb{R}^m$: it is simple to see that this point set has $\lambda_X = m$. These points have the convenient property that each point's projection is independent of each other.

*Proof of Theorem 6.1 from the main paper.* As mentioned previously, the points in $X$ will just be $Re_1, \dots, Re_m$, the $m$ identity unit vectors in $\mathbb{R}^m$ scaled by a factor $R \ge 1$. Since these points each have distance $R\sqrt{2} \ge \sqrt{2}$ from each other, the optimum set of facilities is all of them, which has cost $m$.

Now, consider a random projection $G$ down to $d = o(\log m)$ dimensions, and define $C = \sqrt{\frac{\log n}{10d}}$ and $R = \sqrt{C}$. Note that $R, C = \omega(1)$. Our goal will be to show that with at least $\frac{2}{3}$ probability, for all but $\frac{3m}{C}$ points $p \in GX$, $\tilde{r}_p \le \frac{2}{R}$, where we recall that $\tilde{r}_p$ is the positive real number such that

$$\sum_{q \in B(p, \tilde{r}_p) \cap GX} (\tilde{r}_p - \|p - q\|) = 1.$$

We trivially have the bound $\tilde{r}_p \le 1$ for all $p \in GX$, which means that if we show our goal, then $\sum_{p \in GX} \tilde{r}_p \le m \cdot \frac{2}{R} + \frac{3m}{C} \cdot 1 \le \frac{5m}{R} = o(m)$. However, $\sum_{p \in GX} \tilde{r}_p$ is a constant-factor approximation to the optimum facility location cost by Lemma 3.1 in the main paper, which proves that the facility location cost of $GX$ is $o(m)$.

Now, for each $e_i$, by Equation (1), we have for any $C \ge 6$,

$$\begin{aligned}\Pr(\|Ge_i\| \le C) &\ge 1 - \exp\left(-d(C-1)^2/8\right) \\ &\ge 1 - \exp\left(-(C-1)^2/8\right) \\ &\ge 1 - \frac{1}{2C}.\end{aligned}$$

Moreover, conditioned on $\|Ge_i\| \le C$, by Lemma A.2, we have that for each $j \ne i$, $\Pr(\|Ge_j - Ge_i\| \le \frac{1}{C}) \ge n^{-1/10}$ if $n$ is sufficiently large. Therefore, if $Ge_i : \|Ge_i\| \le C$ is fixed, since the $Ge_j$'s are independent vectors, we can apply the Chernoff bound to say that with probability at least $1 - n^{-10}$, at least $\log n \ge R$ values of $j \ne i$ satisfy $\|Ge_j - Ge_i\| \le \frac{1}{C}$, or equivalently, $\|G(Re_j) - G(Re_i)\| \le \frac{R}{C} = \frac{1}{R}$. By removing our conditioning on $Ge_i$, we have that with probability at least $1 - \frac{1}{2C} - n^{-10} \ge 1 - \frac{1}{C}$, there are at least $R$ points in $GX$ that are within $\frac{1}{R}$ of $G(Re_i)$, in which case we have that for $p = G(Re_i)$, $\tilde{r}_p \le \frac{2}{R}$. Therefore, in expectation, at most $\frac{m}{C}$ of the points in $GX$ have $\tilde{r}_p > \frac{2}{R}$. Thus, by Markov's inequality, with probability at least $\frac{2}{3}$, at most $\frac{3m}{C}$ of the points in $GX$ have $\tilde{r}_p > \frac{2}{R}$. This proves the first part of the theorem.

To prove the second part of the theorem, note that the optimal facility location cost over $GX$ is $o(m)$ with probability at least $2/3$, which implies that the number of open facilities in any optimal solution $\mathcal{F}_d$ is $o(m)$. But then, each point in $X$ which is not an open facility center is at least

$R\sqrt{2}$ away from the nearest open facility center in the original space $X$, so the facility location cost in $X$ is at least $(m - o(m)) \cdot R\sqrt{2} = \omega(m)$. □

Next, we prove Theorems 6.2 and 6.3 (from the main paper). These results prove that the dependence on doubling dimension $d_X$ is required in the projected dimension $d$, both to approximate the cost of the minimum spanning tree and to produce a minimum spanning tree in the lower dimension that is still an approximate MST in the original dimension.

*Proof of Theorem 6.2 from the main paper.* Let $X = \{0, e_1, \dots, e_m\}$, where $m = n - 1$, $0$ is the origin in $\mathbb{R}^m$ and $e_i$ is the $i$th identity vector for each $1 \le i \le m$. Clearly, the minimum spanning tree connects $0$ to all of the $e_i$'s and has cost $M = m$. Now, we show that for $C = \sqrt{\frac{\log n}{10d}} = \omega(1)$, the MST cost of $GX$ is at most $\frac{10m}{C} = o(m)$ for sufficiently large $m$ with at least $2/3$ probability.

To do so, note that since $G$'s entries are independent, $Ge_1, \dots, Ge_m$ are all i.i.d. $\frac{1}{\sqrt{d}} \cdot \mathcal{N}(0, I_d)$. Consider some $e_i, e_j$ and suppose that $\|Ge_i\|, \|Ge_j\| \le C$ but $\|G(e_i - e_j)\| \ge \frac{4}{C}$. Then, if we let $v = G(e_i + e_j)/2$, for each $k \ne i, j$, $\Pr(\|Ge_k - v\| \le \frac{1}{C}) \ge n^{-1/10}$ by Lemma A.2. By the independence of $Ge_1, \dots, Ge_m$, with probability at least $1 - n^{-10}$, there is some $k \ne i, j$ in $[m]$ such that $\|Ge_k - v\| \le \frac{1}{C}$. In this case, the minimum spanning tree of $GX$ would not have the edge $(Ge_i, Ge_j)$, as this edge could be replaced by either the edge $(Ge_i, Ge_k)$ or $(Ge_k, Ge_j)$, both of which are shorter.

Thus, with probability at least $1 - n^{-8}$, if we just connect the points $Ge_i$ over all $i$ with $\|Ge_i\| \le C$ in an MST, every edge has length at most $\frac{4}{C}$. We can create a possibly suboptimal spanning tree by connecting all $Ge_i$ with norm at most $C$ in an MST, connecting one of these vertices arbitrarily to $0 = G \cdot 0$, and finally connecting $Ge_i$ to $0$ for all $i$ with $\|Ge_i\| > C$. The first part has total cost at most $m \cdot \frac{4}{C}$ with probability at least $1 - n^{-8}$. The second part has total cost at most $C$ with probability at least $1 - n^{-8}$ (as long as some $\|Ge_i\| \le C$). Finally, the third part has total expected cost $m \cdot \mathbb{E}[\|Ge_i\| \cdot \mathbb{I}(\|Ge_i\| \ge C)]$, since each edge $e_i$ contributes to the third part only if $\|Ge_i\| \ge C$, and there are $m$ potential vertices $Ge_1, \dots, Ge_m$. However, by the Cauchy-Schwarz inequality, we know that

$$\begin{aligned}
\mathbb{E}\left[\|Ge_i\| \cdot \mathbb{I}(\|Ge_i\| \ge C)\right] &\le \sqrt{\mathbb{E}\left[\|Ge_i\|^2\right] \cdot \Pr(\|Ge_i\| \ge C)} \\
&\le \sqrt{1 \cdot \exp\left(-d \cdot (C-1)^2/8\right)} \\
&\le \exp\left(-(C-1)^2/16\right) \le \frac{1}{C},
\end{aligned}$$

with the final inequality true if $C \ge 7$. Therefore, with probability at least $\frac{4}{5}$, the third part has cost at most $\frac{5m}{C}$ by Markov's inequality. So, with probability at least $\frac{4}{5} - 2n^{-8} \ge \frac{2}{3}$, the total cost of this spanning tree in $GX$ (which may not even be minimal) is at most $\frac{4}{C} \cdot m + C + \frac{5}{C} \cdot m \le \frac{10m}{C}$ assuming $m$ is sufficiently large. □

*Proof of Theorem 6.3 from the main paper.* As in our proof of Theorem 6.2, let $C = \sqrt{\frac{\log n}{10d}} = \omega(1)$. Consider $n = C \cdot m + 1$ and let $X = \{0\} \cup \{e_i \cdot k/C\}$ for $1 \le i \le m, 1 \le k \le C$. The minimum spanning tree connects $0$ to $e_i/C$ to $2e_i/C$ to so on, so each edge has length $1/C$ and the total MST cost is $M = m$.

Now, by Equations (1) and (2), for each $e_i$, the probability that $\|Ge_i\| \in [1/10, 100]$ is at least $1 - \exp(-d/10) - (3/100)^d > 0.06$ for all $d \ge 1$. Thus, with exponential failure probability in $m$, among $e_1, \dots, e_{m/2}$, at least $0.02m$ of the $Ge_i$'s have norm between $1/10$ and $100$. Now, for some $i \le m/2$ with $1/10 \le \|Ge_i\| \le 100$, since $d = o(\log n)$, by Lemma A.2, the probability that $\|Ge_j - Ge_i\| \le \frac{1}{100C}$ for any $j > m/2$ is at least $n^{-1/10}$. Hence, with exponential failure probability, for each $i$ with $1/10 \le \|Ge_i\| \le 100$, there is some $j > m/2$ with $\|Ge_j - Ge_i\| \le \frac{1}{100C}$.

Let $I$ be the set of $i$ such that $\|Ge_i\| \ge \frac{1}{10}$ and there is some $j$ with $\|Ge_j - Ge_i\| \le \frac{1}{100C}$. For each $i \in I$, the distance between $Ge_i \cdot k/C$ and $Ge_i \cdot \ell/C$ for any $\ell \ne k$ is at least $\frac{1}{10C}$ but the distance between $Ge_i \cdot k/C$ and $Ge_j \cdot k/C$ is at most $\frac{1}{100C}$. This means that the closest point to $Ge_i \cdot k/C$ in $GX$ is of the form $Ge_j \cdot k'/C$ for some $j \ne i$ and $k'$ which may or may not equal $k$. However, for every $Gx \in GX$, the minimum spanning tree of $GX$ must contain the edge connecting $Gx$ to its closest neighbor, so for each $i \in I$ and $1 \le k \le C$, $\widetilde{\mathcal{M}}$ must connect $Ge_i \cdot k/C$ to $Ge_j \cdot k'/C$, which has length at least $k/C$ in the original space $\mathbb{R}^m$. Therefore, the pullback of the MST has length at least

$$\sum_{i \in I} \sum_{k=1}^{C} \frac{k}{C} \ge \frac{C}{2} \cdot |I|,$$

which with exponential failure probability in $m$ is at least $\frac{C}{100} \cdot m = \frac{C}{100} \cdot M = \omega(M)$. □

### E.2. Approximate Solutions Cannot be Pulled Back

In this subsection, we prove Lemmas 6.4 and 6.5 from the main paper. In other words, we give a simple example showing that our definition of *locally optimal* (for FL) and that *optimal* (for MST) is necessary, if we want dependence on $d_X = \log \lambda_X$ as opposed to $\log n$. In particular, our lemmas give examples showing that pulling back of *any* approximately optimal solution found in the projected space to the original space does not work.

*Proof of Lemma 6.4 from the main paper.* Consider the following set of points $Y$:

$$Y = \{b_1, b_2, \dots, b_m\} = \{e_1, e_1+e_2, \dots, e_1+e_2+\cdots+e_m\}$$

where $e_i$ is the $i$th standard basis vector. We refer to this dataset as the 'walk' dataset. Using the definition of doubling dimension (see Section 2 in the main paper), we can compute that the doubling dimension of $Y$ is some constant independent of $m$. Now construct the dataset $X$ by scaling all the points in $Y$ by the factor $m^{1+1/2d}$. This does not affect the doubling dimension. Consider the projection of $X$ into $\mathbb{R}^d$ where $d = O(1)$. Before projection, the optimum solution is to open all facilities, costing $m$.

Now consider applying a random projection $G$ and note that the projection of the differences $G(b_i - b_{i+1})$ are independent. Therefore, by Proposition A.1, there is a pair of consecutive points $b_i, b_{i+1}$ such that $\|G(b_i - b_{i+1})\|$ shrinks by a factor of $C_1/m^{1/d}$ with probability at least $9/10$. Furthermore, by Equation (2), we have that all the differences $\|G(b_i - b_{i+1})\|$ do not shrink by a factor worse than $C_2/m^{1/d}$ with probability at least $9/10$. Hence, with some constant probability, *both* the following events occur:

- There exists some $i^*$ such that $\|G(b_{i^*} - b_{i^*+1})\| = O(m^{1-1/2d})$
- $\|G(b_i - b_{i+1})\| = \Omega(m^{1-1/2d})$ for all $i$.

In this case, the optimal solution in the projected space is to include all facilities, which has total cost $m$. However, a solution that is within a $1 + O(m^{-1/2d})$ multiplicative factor of the optimal solution is to include all facilities except for $Gb_i^*$. However, evaluating this solution in the original dimension incurs a cost at least $\Omega(m^{1+1/2d})$, whereas the optimal cost is still $m$. Hence, the approach has approximation ratio of at least $m^{1/2d}$, which is $\omega(1)$, i.e., superconstant unless $d = \Omega(\log m)$. □

*Proof of Lemma 6.5 from the main paper.* Assume WLOG that $n = 2k^2$ for some $k$, that $X$ lies in $\mathbb{R}^m$ for $m = k+1$, and that $d = \epsilon \cdot \log n$ for some $\epsilon = o(1)$. Now, let $e_1, e_2, \dots, e_k$ represent the identity vectors in $\mathbb{R}^k$. Now, we will choose our $n$ points as follows. First, we will choose the $k^2$ points $X' = \{(0, \mathbf{0}), (\frac{1}{k}, \mathbf{0}), \dots, (\frac{k^2-1}{k}, \mathbf{0})\}$, where $\mathbf{0}$ represents the last $k$ coordinates all being 0. For the remaining $k^2$ points, for each $0 \le i \le k-1$ we add the set $X_i = \{(i, e_i), (i + \frac{1}{k}, e_i) \dots, (i + \frac{k-1}{k}, e_i)\}$. We let $X = X' \cup X_0 \cup \cdots \cup X_{k-1}$.

First, we show that the doubling dimension of $X$, $\lambda_X$, is at most $O(1)$. First, note that $X'$ and each $X_i$ is trivially embeddable into one dimension, because the points in $X'$ and in each $X_i$ only vary on one coordinate, so each of these individually have doubling dimension $O(1)$. Therefore, for any ball $B = B(r, p)$ of radius $r \le 10$ around some point $p$, $B \cap X$ is contained in some union of $O(1)$ of $X', X_0, \dots, X_{k-1}$. Consequently, the points in $B \cap X$ can be decomposed into $O(1)$ balls of radius $r/2$, since $B \cap X'$ and $B \cap X_i$ each have doubling dimension bounded by a constant. Now, if we consider some ball $B = B(r, p)$ of radius $r > 10$, suppose that $p = (a_0, a_1, \dots, a_k) \in \mathbb{R}^{k+1}$. Now, consider the 5 points $\{(a_0 + \frac{j}{2} \cdot r, \mathbf{0})\}_{j=-2}^{2}$, where the $\mathbf{0}$ represents the last $k$ coordinates all being 0. For every point $x$ in $X \cap B$, $x$'s first coordinate must be in the range $[a_0 - r, a_0 + r]$ and $x$'s remaining coordinates have total magnitude at most 1. With these two observations, it is immediate that every point in $X \cap B$ is within $r/2$ of some point $\{(a_0 + \frac{j}{2} \cdot r, \mathbf{0})\}$ for some integer $-2 \le j \le 2$. Therefore, if $r > 10$, $B \cap X$ can be covered by 5 balls of radius $r/2$. Thus, $\lambda_X = O(1)$, so $X$ has doubling dimension $\log \lambda_X = O(1)$.

Now, a straightforward verification tells us that for any $i \ne j$, the points in $X_i$ and the points in $X_j$ are at least $\sqrt{2}$ away from each other. Moreover, each point $(i + \frac{j}{k}, e_i)$'s closest point in $X'$ is the corresponding point $(i + \frac{j}{k}, \mathbf{0})$, and this distance is 1. Therefore, the minimum spanning trees of $X$ are as follows. First, connect the points in $X'$ in a line and all of the points in each $X_i$ in a line. Finally, for each $0 \le i \le k-1$, choose some arbitrary $j$ and connect $(i + \frac{j}{k}, e_i)$ and $(i + \frac{j}{k}, \mathbf{0})$. The total MST cost $M$ is $\frac{k^2-1}{k} + k \cdot \frac{k-1}{k} + k \cdot 1 = 3k - 1 - \frac{1}{k} = (3 - o(1))k$.

Now, when the random projection $G : \mathbb{R}^{k+1} \to \mathbb{R}^d$ is applied, we have that each vector $(0, e_i)$ is independently mapped to some vector $(a_{i1}, \dots, a_{id})$, where each $a_{ij}$ for $1 \le i \le k, 1 \le j \le d$ is an i.i.d. $\mathcal{N}(0, 1/d)$. So for any $\epsilon = o(1)$ and $n$ sufficiently large, if we choose $\delta = e^{-1/(100\epsilon)}$, we have that $\Pr(|a_{i1}|, \dots, |a_{id}| \le \delta/\sqrt{d}) = \Theta(\delta)^d \le e^{-\log n/4} < 1/\sqrt{2k}$, where we used the fact that $d = \varepsilon \log n$. Hence, a simple Chernoff bound tells us that with $1 - o(1)$ probability, at least $\sqrt{k}/2$ of the $(0, e_i)$'s get mapped to some $(a_{i1}, \dots, a_{id})$ with norm at most $\delta$.

Now, consider the following $\omega(1)$-approximate MST for $X$. Let $A = \epsilon^{-1}$, and choose some set $I = \{i_1, \dots, i_A\}$. Our "approximate" MST will be as follows. For each $i \in I$, remove the $k-1$ edges connecting $X_i$ together, and for each $1 \le j \le k$, connect $(i + \frac{j}{k}, 0)$ with $(i + \frac{j}{k}, e_i)$. Each time this is done, we remove $k-1$ edges of length $1/k$ and add $k-1$ edges of length 1 (recall that one of these edges of length 1 was already in the MST), so the MST cost increases by $\epsilon^{-1}((k-1)1 - (k-1)/k) = \epsilon^{-1} k \cdot (1 - o(1))$. Hence, regardless of what set $A$ we chose, the approximate MST is a $\omega(1)$-approximation, as the true MST has cost $M = O(k)$.

However, we claim that with high probability, we can choose $A$ so that this becomes a $(1 + o(1))$-approximation in the projected space. Indeed, since $\epsilon \ge \frac{1}{\log n}$, with $1 - o(1)$

probability, at least $\sqrt{k}/2 \geq \epsilon^{-1}$ values $e_i$ get mapped to some point with norm at most $\delta$. So, we choose $A$ to be of size $\epsilon^{-1}$ so that for all $i \in A$, $e_i$ gets mapped to a point with norm at most $\delta$. Recall that $\mathcal{M}$ denote the true MST for $X$, and let $\mathcal{M}'$ be this poor-approximation spanning tree. Note that the only edges in $\mathcal{M}'\backslash\mathcal{M}$ connect $(i + \frac{j}{k}, 0)$ to $(i + \frac{j}{k}, e_i)$ for $i \in I, 0 \leq j \leq k-1$. Since there are $\varepsilon^{-1} \cdot k$ such edges, and each edge has size at most $\delta$ when projected, we have that

$$\begin{aligned}\text{cost}_{GX}(\mathcal{M}') &\leq \text{cost}_{GX}(\mathcal{M}) + \delta \cdot \varepsilon^{-1} \cdot k \\ &\leq \text{cost}_{GX}(\mathcal{M}) + \varepsilon^{-1} \cdot e^{-\varepsilon^{-1}/100} \cdot k \\ &= \text{cost}_{GX}(\mathcal{M}) + o(k).\end{aligned}$$

Now, let's suppose that $d \geq \omega(\log \log n)$. We saw in subsection D.1 that $\text{cost}_{GX}(\mathcal{M})$ had expectation at most $M = \text{cost}_X(\mathcal{M})$ and standard deviation $O(M/\sqrt{\log \log n})$, regardless of the dataset $X$. So, with $9/10$ probability, $\text{cost}_{GX}(\mathcal{M}') \leq \text{cost}_{GX}(\mathcal{M}) + o(k) = (1+o(1))M$. Moreover, by Theorem 5.1 from the main paper, with $9/10$ probability, $\widetilde{M}$, the cost of the MST in the reduced space $GX$, is within a $1 \pm o(1)$ factor of $M$. Therefore, with at least $4/5 - o(1)$ probability, $\text{cost}_{GX}(\mathcal{M}') \leq (1 + o(1)) \cdot \widetilde{M}$, so $\mathcal{M}'$ is an $\omega(1)$-approximate MST in $X$ but a $1 + o(1)$-approximate MST in $GX$. □

### E.3. Lower Bounds for $k$-means and $k$-medians

In this subsection, we prove Theorem 6.6 from the main paper, which shows the tightness of the bounds of (Makarychev et al., 2019) for $k$-means and $k$-medians clustering even in the case of *constant doubling dimension*.

We remark that (Makarychev et al., 2019) showed tightness of their result if doubling dimension is ignored. Namely, they showed the existence of such a point set $X$ that may have large doubling dimension. Hence, our contribution is making such a set that also has doubling dimension $O(1)$.

*Proof of Theorem 6.6 from the main paper.* We start with the case where $n = 2t$ and $k = 2t - 1$ for some $t$. As in (Makarychev et al., 2019), we wish to consider $t$ pairs of points where each pair is of distance 1 from each other, but all other distances are larger.

Namely, we do the following. First, define $D = t^{1/d}/10$, and let $R = \sqrt{D}$. We have that $D, R = \omega(1)$, since $d = o(\log n) = o(\log t)$. Now, for $1 \leq i \leq t$, let $a_i = (2 \cdot i, \mathbf{0})$, meaning that $a_i$'s first coordinate is $2 \cdot i$ and the remaining $t = m - 1$ coordinates are 0. Next, for each $1 \leq i \leq t - 1$, define $b_i = a_i + e_{i+1}$, i.e., $b_i$ has first coordinate $2 \cdot i$, $(i + 1)$th coordinate 1, and all remaining coordinates 0. However, define $b_t = a_t + \frac{1}{R} \cdot e_{i+1}$. Our set $X$ will be the union of the $a_i$'s and $b_i$'s.

Now, since $k = n - 1$, the $k$-medians cost of $X$ is just the distance between the closest pair of points in $X$, which is $\frac{1}{R}$. The $k$-means cost of $X$ is just the squared distance between the closest pair of points in $X$. However, by Proposition A.1, for each $i$,

$$\begin{aligned}\Pr\left(\|Gb_i - Ga_i\| \leq \frac{10}{t^{1/d}}\right) &= \Pr\left(\|Ge_{i+1}\| \leq \frac{10}{t^{1/d}}\right) \\ &\geq \left(\frac{10}{e \cdot t^{1/d}}\right)^d \geq \frac{3}{t}.\end{aligned}$$

Moreover, since $e_2, \dots, e_t$ are all distinct unit vectors, the vectors $Ge_2, \dots, Ge_t$ are independent, which means that with probability at least $1 - (1 - 3/t)^{t-1} \geq 0.9$ (for $t$ sufficiently large), some $1 \leq i \leq t - 1$ will have $\|Gb_i - Ga_i\| \leq 10/t^{1/d} = 1/D$. Thus, some pair of points $(a_i, b_i)$ satisfy $\|Ga_i - Gb_i\| \leq 1/D$, whereas the closest distance between two points in $X$ was only $1/R$. Therefore, with at least $9/10$ probability, the $k$-medians cost has multiplied by a $R/D = o(1)$ factor after projection, and likewise, the $k$-means cost has multiplied by a $R^2/D^2 = o(1)$ factor.

Now, let $p, q \in X$ be the pair of points minimizing $\|Gp - Gq\|$. With probability at least $4/5$, $\|Ga_t - Gb_t\| \geq 1/(20R) > 1/D$, which means that either $p$ or $q$ is not in $\{a_t, b_t\}$: assume WLOG that $p \notin \{a_t, b_t\}$. Thus, an optimal choice of $k$ centers (for either $k$-means or $k$-medians) is choosing all points in $X$, except $p$. But then, in the original space, these centers have $k$-medians cost equal to the distance from $p$ to its closest point in $X$, which is at least 1. Likewise, the $k$-means cost is also at least 1. However, the optimal $k$-medians and $k$-means costs are $1/R$ and $1/R^2$, respectively, so the optimal choice in $GX$ is an $R = \omega(1)$ or $R^2 = \omega(1)$ approximation for $k$-medians and $k$-means, respectively. This finishes the proof in the case that $k = n - 1$.

For general values of $k < n$, we can simply consider having $n' = k + 1$ points in the configuration as above, but with exactly one of the points replicated $n - k$ times. In this case, the cost of $k$-medians clustering is still the distance of the closest pair of distinct points, and the cost of $k$-medians clustering is still the square of the distance of the closest pair of distinct points. So, the lower bound of $\Omega(\log k)$ still holds. □

## F. Facility Location with Squared Costs

Recall that the facility location with squared costs problem is defined as follows. Given a dataset $X \subset \mathbb{R}^m$, our goal is to find a subset $\mathcal{F} \subseteq X$ that minimizes the objective

$$\text{cost}(\mathcal{F}) = |\mathcal{F}| + \sum_{x \in X} \min_{f \in \mathcal{F}} \|x - f\|^2. \quad (25)$$

Similar to Equation (3) in the main paper, we give a geometric expression that is a constant factor approximation to the cost of the objective presented in (25). For each $p \in X$, associate it with a radius $r_p > 0$ that satisfies the relation

$$\sum_{q \in B(p,r)} (r_p^2 - \|p - q\|^2) = 1. \tag{26}$$

We generalize the results in (Mettu & Plaxton, 2000) and (Badoiu et al., 2005) to give an analogue of Lemma 3.1 in the main paper for the squared objective (25).

**Lemma F.1.** *Let $C_{OPT}$ denote the cost of the optimal solution to the objective given in* (25)*. Then*

$$\frac{1}{8} \cdot C_{OPT} \leq \sum_{p \in X} r_p^2 \leq 24 \cdot C_{OPT}.$$

To prove Lemma F.1, we first given an algorithm for (25) inspired by the MP algorithm. Our algorithm, which we denote as the 'Squared MP Algorithm,' is the following.

---

**Algorithm 2** SQUARED MP ALGORITHM

**Input** : Dataset $X = \{p_1, \cdots, p_n\} \subseteq \mathbb{R}^d$
**Output** : Set $\mathcal{F}$ of facilities

7 $\mathcal{F} \leftarrow \emptyset$ **for** $i = 1$ *to* $n$ **do**
8     Compute $r_i$ satisfying: $\sum_{q \in B(p_i, r_i)} (r_i^2 - \|p_i - q\|^2) = 1$
9 Sort such that $r_1 \leq \ldots \leq r_n$ **for** $i = 1$ *to* $n$ **do**
10     **if** $B(p_i, 2r_i) \cap \mathcal{F} = \emptyset$ **then**
11         $\mathcal{F} \leftarrow \mathcal{F} \cup \{p_i\}$
12 Output $\mathcal{F}$

---

We first claim that the set of facilities returned by Algorithm 2 is a constant factor approximation to the optimal set.

**Theorem F.2.** *Let $C_{OPT}$ denote the cost of the optimal solution to the objective given in* (25) *and let $\mathcal{F}$ denote the set of facilities returned by Algorithm* 2*. Then* $\text{cost}(\mathcal{F}) \leq 6 \cdot C_{OPT}$.

*Proof.* The proof follows similarly to Theorem 1 in (Mettu & Plaxton, 2000) with some adaptations. Let $\mathcal{F}'$ denote any set of facilities. For any point $x \in X$, let

$$\text{charge}(x, \mathcal{F}') = d(x, \mathcal{F}')^2 + \sum_{p \in \mathcal{F}'} \max(0, r_p^2 - \|p - x\|^2)$$

where $d(x, \mathcal{F}')$ denotes the distance between $x$ and the closest point to $x$ in $\mathcal{F}'$ and $r_p$ is defined as in (26). We first show that $\sum_{x \in X} \text{charge}(x, \mathcal{F}') = \text{cost}(F')$. Indeed, this follows from swapping the order of summation:

$$\begin{aligned}
&\sum_{x \in X} \text{charge}(x, \mathcal{F}') \\
&= \sum_{x \in X} \sum_{p \in \mathcal{F}'} \max(0, r_p^2 - \|p - x\|^2) + \sum_{x \in X} d(x, \mathcal{F}')^2 \\
&= \sum_{p \in \mathcal{F}'} \sum_{x \in X} \max(0, r_p^2 - \|p - x\|^2) + \sum_{x \in X} d(x, \mathcal{F}')^2 \\
&= \sum_{p \in \mathcal{F}'} 1 + \sum_{x \in X} d(x, \mathcal{F}')^2 = \text{cost}(\mathcal{F}').
\end{aligned}$$

Now denote $F^*$ as the set of facilities for the optimal solution. We first study the individual term $\text{charge}(x, \mathcal{F}^*)$. We first give a lower bound for $\text{charge}(x, \mathcal{F}^*)$. Let $q^*$ be the closest point to $x \in \mathcal{F}^*$. If $x \notin B(q^*, r_{q^*})$ then $\text{charge}(x, \mathcal{F}^*) \geq \|x - q^*\|^2 > r_{q^*}^2$. Otherwise,

$$\begin{aligned}
\text{charge}(x, \mathcal{F}^*) &\geq \|x - q^*\|^2 + r_{q^*}^2 - \|x - q^*\|^2 \\
&= r_{q^*}^2 \geq \|x - q^*\|^2
\end{aligned}$$

so altogether,

$$\text{charge}(x, \mathcal{F}^*) \geq \max(r_{q^*}^2, \|x - q^*\|^2). \tag{27}$$

Now let $\mathcal{F}$ denote the set of solutions returned by Algorithm 2. We now upper bound $\text{charge}(x, \mathcal{F})$ in terms of the quantities $r_{q^*}^2, \|x - q^*\|^2$. Recall that $q^* \in \mathcal{F}^*$ is the closest point to $x$ in $\mathcal{F}^*$. We note that there must be a point $q \in \mathcal{F}$ such that $r_q \leq r_{q^*}$ and $\|q - q^*\| \leq 2r_{q*}$ due to how Algorithm 2 selects the set of facilities in step 6.

Now if $x \in B(q, r_q)$ then $d(x, \mathcal{F}) \leq \|x - q\|$ and thus $\text{charge}(x, \mathcal{F}) \leq r_q^2$ since step 6 of Algorithm 2 insures that $x \notin B(q', r_{q'})$ for any other $q' \in \mathcal{F}$. Otherwise, $x \notin B(q, r_q)$ in which case we claim that $\text{charge}(x, \mathcal{F}) \leq \|x - q\|^2$. This claim is immediate unless there exists some $q' \in \mathcal{F}$ such that $x \in B(q', r_{q'})$. However in this case, a similar reasoning as above means $\text{charge}(x, \mathcal{F}) \leq r_{q'}^2$ but

$$\|x - q\| \geq \|q - q'\| - \|x - q'\| > 2r_{q'} - r_{q'} = r_{q'}$$

where the second inequality again follows from step 6 of Algorithm 2. Therefore,

$$\begin{aligned}
\text{charge}(x, \mathcal{F}) \leq \|x - q\|^2 &\leq (\|x - q^*\| + \|q^* - q\|)^2 \\
&\leq 2\|x - q^*\|^2 + 2\|q^* - q\|^2 \\
&\leq 2\|x - q^*\|^2 + 4r_{q^*}^2.
\end{aligned} \tag{28}$$

Comparing (27) to (28), we can compute that the ratio of $2\|x - q^*\|^2 + 4r_{q^*}^2$ to $\max(r_{q^*}^2, \|x - q^*\|^2)$ is at most 6 from which it follows that

$$\text{charge}(x, \mathcal{F}) \leq 6 \cdot \text{charge}(x, \mathcal{F}^*).$$

Summing over $x \in X$ completes the proof. □

Using Theorem F.2, we are now in position to prove Lemma F.1. The proof of Lemma F.1 follows similarly to the proof of Lemma 2 in (Badoiu et al., 2005) with some modifications to suit our alternate objective function given in (25).

*Proof of Lemma F.1.* We first prove the lower bound. Note that for every $p_i \in X$, Algorithm 2 will open a facility within distance at most $2r_p$. Hence, $4\sum_{p\in X} r_p^2$ is an upper bound on the cost to connect the points to their nearest facility. Now from similar reasoning as in the proof of Theorem F.2, we note that each $p$ is in at most one ball $B(q, r_q)$ for some $q \in \mathcal{F}$, where $\mathcal{F}$ denotes the set of facilities returned by Algorithm 2. Therefore,

$$\sum_{p\in X} r_p^2 \geq \sum_{q\in\mathcal{F}} \sum_{p\in B(q,r_q)} r_p^2.$$

Now if $p \in B(q, r_q)$ for some $q \in \mathcal{F}$ then we must have $r_q \leq 2r_p$ because otherwise, step 6 of Algorithm 2 would not have chosen $q$ as a facility center. Thus,

$$\sum_{p\in X} r_p^2 \geq \sum_{q\in\mathcal{F}} \sum_{p\in B(q,r_q)} r_p^2 \geq \frac{1}{4}\sum_{q\in\mathcal{F}} r_q^2 \cdot |B(q, r_q)|.$$

Finally, we know that

$$1 = \sum_{p\in B(q,r_q)} (r_q^2 - \|p-q\|^2) \leq r_q^2 \cdot |B(q, r_q)|$$

from which it follows that $4\sum_{p\in X} r_p^2 \geq |\mathcal{F}|$. Altogether, we see that $8\sum_{p\in X} r_p^2$ is an upper bound to the cost of the solution returned by Algorithm 2 so the lower bound follows.

For the upper bound, we will show that the sum of the radii squared is not too large compared to $\mathrm{cost}(\mathcal{F})$ where $\mathcal{F}$ is the set of facilities returned by Algorithm 2. Consider $p \notin \mathcal{F}$ and let $q$ be the closest facility to $p$. First, we must have $r_p^2 \leq 2(\|p-q\|^2 + r_q^2)$ because otherwise, $r_p^2 > (\|p-q\| + r_q)^2$ which implies that $B(q, r_q) \subseteq B(p, r_p)$. Furthermore,

$$\begin{aligned}
&\sum_{p'\in B(p,r_p)} (r_p^2 - \|p-p'\|^2)\\
\geq &\sum_{p'\in B(q,r_q)} (r_p^2 - \|p-p'\|^2)\\
> &\sum_{p'\in B(q,r_q)} (2r_q^2 + 2\|p-q\|^2 - \|p-p'\|^2)\\
\geq &\sum_{p'\in B(q,r_q)} (r_q^2 + 2\|p-q\|^2 + \|p'-q\|^2 - \|p-p'\|^2)\\
\geq &\sum_{p'\in B(q,r_q)} (r_q^2 - \|q-p'\|^2) = 1
\end{aligned}$$

which contradicts (26). To summarize, if $p \notin \mathcal{F}$ and $q$ is the closest facility in $\mathcal{F}$ to $p$, then

$$r_p^2 \leq 2(\|p-q\|^2 + r_q^2). \tag{29}$$

Going back to the upper bound, recall the definition of $\mathrm{charge}(p, \mathcal{F})$ used in the proof of Theorem F.2:

$$\mathrm{charge}(p, \mathcal{F}) = d(p, \mathcal{F})^2 + \sum_{q\in\mathcal{F}} \max(0, r_q^2 - \|q-p\|^2).$$

We also showed there that $\sum_{p\in X} \mathrm{charge}(p, \mathcal{F}) = \mathrm{cost}(\mathcal{F})$. Now

$$\begin{aligned}
\mathrm{cost}(\mathcal{F}) &= \sum_{p\in X} \mathrm{charge}(p, \mathcal{F})\\
&\geq \sum_{q\in\mathcal{F}} r_q^2 + \sum_{p\in X\setminus\mathcal{F}} \max(r_{\delta(p)}^2, \|p-\delta(p)\|^2)
\end{aligned}$$

where $\delta(p)$ denotes the closest element in $\mathcal{F}$ to $p$. From (29), we know that $r_p^2 \leq 2(\|p-q\|^2 + r_q^2)$ so $\max(r_{\delta(p)}^2, \|p-\delta(p)\|^2) \geq r_p^2/4$ which gives us

$$6 \cdot C_{OPT} \geq \mathrm{cost}(\mathcal{F}) \geq \frac{1}{4} \cdot \sum_{p\in X} r_p^2,$$

as desired. □

We can prove the following statements about the expected value of $r_p$, defined as in (26), after a random projection to a suitable dimension depending on the doubling dimension of the set $X$. The following lemma is analogous to Lemmas C.2 and C.1 and omit its proof since the proof follows identically from the proofs in Lemmas C.2 and C.1.

**Lemma F.3.** *Let $X \subseteq \mathbb{R}^m$ and let $p \in X$. Let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X)$. Let $r_p$ and $\tilde{r}_p$ be the radius of $p$ and $Gp$ in $\mathbb{R}^m$ and $\mathbb{R}^d$ respectively, computed according to Eq. (26). Then there exist constants $c, C > 0$ such that*

$$cr_p^2 \leq \mathbb{E}[\tilde{r}_p^2] \leq Cr_p^2.$$

Combining Lemma F.3, which states that $\sum_p r_p^2$ is a constant factor approximation to thelb optimal solution of the objective given in (25), with Lemma F.1, we obtain the following theorem that is analogous to Theorem 4.1 in the main paper.

**Theorem F.4.** *Let $X \subseteq \mathbb{R}^m$ and let $p \in X$. Let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X)$. Let $\mathcal{F}_m$ be the optimal solution in $\mathbb{R}^m$ and let $\mathcal{F}_d$ be the optimal solution for the dataset $GX \subseteq \mathbb{R}^d$. Then there exists constants $c, C > 0$ such that*

$$c \cdot \mathrm{cost}(\mathcal{F}_m) \leq \mathbb{E}[\mathrm{cost}(\mathcal{F}_d)] \leq C \cdot \mathrm{cost}(\mathcal{F}_m).$$

Note that the crucial ingredient in the proof of Theorem C.4 that allowed us to connect properties of the doubling dimension to facility location clustering was the relation given in Equation (3) in the main paper. The analogous relation for our new objective function in (25) is given in (26) and one can easily check that the steps in the proof of Theorem C.4 transfer. Therefore, we have the following theorem.

**Theorem F.5.** *Let $X \subseteq \mathbb{R}^m$ and let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X \cdot \log(1/\epsilon)/\epsilon^2)$. Fix $p \in X$ and let $Gx$ be any point in $B(Gp, C\tilde{r}_p)$ in $\mathbb{R}^d$ where $C$ is a fixed constant and $\tilde{r}_p$ is computed according to Eq. (26) in $\mathbb{R}^d$. Then*

$$\mathbb{E}\|p - x\| \leq 2C(1 + O(\epsilon))r_p.$$

To derive a statement analogous to Theorem 4.2 from the main paper for our alternate objective function, we need a notion of a locally optimal solution. This task also follows from using Section 3 of the main paper as a blue print. In particular, we can define local optimality of a solution to (25) as follows.

**Definition F.6.** *A solution $\mathcal{F}$ to the objective given in (25) is* locally optimal *if for all $p \in X$, we have $B(p, 3r_p) \cap \mathcal{F} \neq \emptyset$ where $r_p$ is computed as in (26).*

Then the following lemma follows similarly to Lemma 3.3 of the main paper.

**Lemma F.7.** *Let $\mathcal{F}$ be an any collection of facilities. If there exists a $p \in X$ such that $B(p, 3r_p) \cap \mathcal{F} = \emptyset$, then $\text{cost}(\mathcal{F} \cup \{p\}) < \text{cost}(\mathcal{F})$, i.e., we can improve the solution.*

Finally, as a corollary to Lemma F.7 and Theorem F.5, we have the following corollary.

**Corollary F.8.** *Let $X \subset \mathbb{R}^m$ and let $G$ be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$ for $d = O(\log \lambda_X \cdot \log(1/\epsilon)/\epsilon^2)$. Let $\mathcal{F}_d$ be a locally optimal solution for the dataset $GX$ for the objective function given in (25). Then, the cost of $\mathcal{F}_d$ evaluated in $\mathbb{R}^m$, denoted as $\text{cost}_m(\mathcal{F}_d)$, satisfies*

$$\mathbb{E}[\text{cost}_m(\mathcal{F}_d)] \leq |\mathcal{F}_d| + C' \cdot \sum_{p \in X} r_p$$

*for some constant $C' > 0$.*

**Remark F.9.** *We can compute that a constant smaller than 3 works for Definition F.6 and consequently Lemma F.7 but this choice is inconsequential since we already incur a multiplicative constant factor in Theorem F.5.*

Finally, we argue that the lower bound of Theorem 6.1 in the main paper also carries over to our new objective function, meaning that the dimension we project to must depend on the doubling dimension. We define the connection cost of the objective (25) as the second portion.

**Theorem F.10.** *Let $d = o(\log n)$ and let $G$ be be a random projection from $\mathbb{R}^m$ to $\mathbb{R}^d$. There exists $X \subseteq \mathbb{R}^m$ where $|X| = n$ such that with at least $2/3$ probability, the optimal cost multiplies by $o(1)$ when projected. In addition, there exists an optimal solution $\widetilde{\mathcal{F}}$ in $\mathbb{R}^d$ that is only an $\omega(1)$-approximate solution in the original space $\mathbb{R}^m$.*

*Proof Sketch.* The proof follows similarly as in the proof of Theorem 6.1 in the main paper. We again define $X = \{Re_1, \ldots, Re_m\}$, where $R = \sqrt{C}$ and $C = \sqrt{\frac{\log n}{10d}}$. As in the proof of Theorem 6.1, we again have for any fixed $p = Re_i$, with probability at least $1 - \frac{1}{C}$, there are at least $R$ points in $GX$ within $\frac{1}{R}$ distance of $Gp$. For any such point $p$, letting $\tilde{r}_p$ be the associated radius for $GX$ around $Gp$ as computed by Equation (25), we have that $\tilde{r}_p \leq \frac{2}{\sqrt{R}}$. So, with at least $2/3$ probability, at most $\frac{3m}{C}$ of the points have $\tilde{r}_p > \frac{2}{\sqrt{R}} = o(1)$. As in the proof of Theorem 6.1, this shows that the optimal cost multiplies by a $o(1)$ factor, by using Lemma F.1 this time.

In the original space $X \subset \mathbb{R}^m$, the optimal squared facility location cost is $m$, which is achievable by setting every point in $X$ as a facility. However, since the optimal facility cost in $GX$ is $o(m)$, the optimal solution $\widetilde{\mathcal{F}}$ in the reduced space $\mathbb{R}^d$ assigns at most $o(m)$ points to be facilities. Therefore, for the remaining $m - o(m)$ points, the connection cost in the original space is at least $(R\sqrt{2})^2 \geq R^2$, so the cost of $\widetilde{\mathcal{F}}$ in the original space $X$ is at least $R^2 \cdot (m - o(m)) = \omega(1) \cdot m$. Thus, any optimal solution $\widetilde{\mathcal{F}}$ is an $\omega(1)$-approximate solution in the original space $\mathbb{R}^m$. $\square$

## References


Badoiu, M., Czumaj, A., Indyk, P., and Sohler, C. Facility location in sublinear time. In Caires, L., Italiano, G. F., Monteiro, L., Palamidessi, C., and Yung, M. (eds.), *Automata, Languages and Programming*, pp. 866–877, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg. ISBN 978-3-540-31691-6.

Indyk, P. and Naor, A. Nearest-neighbor-preserving embeddings. *ACM Trans. Algorithms*, 3(3):31–es, August 2007. ISSN 1549-6325. doi: 10.1145/1273340.1273347. URL https://doi.org/10.1145/1273340.1273347.

Kou, L. T., Markowsky, G., and Berman, L. A fast algorithm for steiner trees. *Acta Informatica*, 15:141–145, 1981. doi: 10.1007/BF00288961. URL https://doi.org/10.1007/BF00288961.

Makarychev, K., Makarychev, Y., and Razenshteyn, I. Performance of johnson-lindenstrauss transform for k-means and k-medians clustering. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pp. 1027–1038, New York, NY, USA, 2019. Association for Computing

Machinery. ISBN 9781450367059. doi: 10.1145/3313276.3316350. URL https://doi.org/10.1145/3313276.3316350.

Mettu, R. R. and Plaxton, C. G. The online median problem. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pp. 339–348, Nov 2000. doi: 10.1109/SFCS.2000.892122.

Wolfram Research. Chi distribution. URL https://mathworld.wolfram.com/ChiDistribution.html.