Metric 1-median selection with fewer queries

Ching-Lueh Chang

December 28, 2016

Abstract

Let \( h: \mathbb{Z}^+ \rightarrow \mathbb{Z}^+ \setminus \{1\} \) be such that (1) \( h(n) \leq \lg n \) for all sufficiently large \( n \) and (2) \( h(n) \) and \( \lceil n^{1/h(n)} \rceil \) are computable from \( n \) in \( O(h(n) \cdot n^{1+1/h(n)}) \) time. We show that given an \( n \)-point metric space \((M, d)\), the problem of finding \( \min_{i \in M} \sum_{j \in M} d(i, j) \) (breaking ties arbitrarily) has a deterministic, \( O(h(n) \cdot n^{1+1/h(n)}) \)-time, \( O(n^{1+1/h(n)}) \)-query, \( (2h(n)) \)-approximation and nonadaptive algorithm. Our proofs modify those of Chang [2, 3].

1 Introduction

A metric space is a nonempty set \( M \) endowed with a function \( d: M \times M \rightarrow [0, \infty) \) such that

\[
\begin{align*}
  d(x, x) &= 0, \\
  d(x, y) &> 0, \\
  d(x, y) &= d(y, x), \\
  d(x, y) + d(y, z) &\geq d(x, z)
\end{align*}
\]

for all distinct \( x, y, z \in M \) [11]. Given an \( n \)-point metric space \((\{0, 1, \ldots, n - 1\}, d)\), METRIC 1-MEDIAN asks for \( \min_{i=0}^{n-1} \sum_{j=0}^{n-1} d(i, j) \), breaking ties arbitrarily. It has a Monte-Carlo \( O(n/\epsilon^2) \)-time \((1 + \epsilon)\)-approximation algorithm for all constants \( \epsilon > 0 \) [8, 9]. Kumar et al. [10] give a Monte-Carlo \( O(D \cdot \exp(1/\epsilon^{O(1)})) \)-time \((1 + \epsilon)\)-approximation algorithm for 1-median selection in \( \mathbb{R}^D \), where \( \epsilon > 0 \) and \( D \in \mathbb{Z}^+ \). Algorithms abound for the more general metric \( k \)-median problem [6, 7, 10].

This paper focuses on deterministic sublinear-time algorithms for METRIC 1-MEDIAN, where “sublinear” means “\( o(n^2) \)” because there are \( n(n-1)/2 \) nonzero distances. In particular, we shall improve the following theorem.

\*Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan.
Email: clchang@saturn.yzu.edu.tw
**Theorem 1** (Implicit in [1, 2, 12]). Let $h: \mathbb{Z}^+ \to \mathbb{Z}^+-\{1\}$ be such that (1) $h(n) \leq \lg n$ for all sufficiently large $n$ and (2) $h(n)$ and $[n^{1/h(n)}]$ are computable from $n$ in $O(h(n) \cdot n^{1+1/h(n)})$ time. Then METRIC 1-MEDIAN has a deterministic, $O(h(n) \cdot n^{1+1/h(n)})$-time, $O(h(n) \cdot n^{1+1/h(n)})$-query, $(2h(n))$-approximation and nonadaptive algorithm.

To prove Theorem 1, Chang [2] designs a function $\tilde{d}: \{0, 1, \ldots, n - 1\}^2 \to [0, \infty)$ such that a 1-median w.r.t. $\tilde{d}$ is $(2h(n))$-approximate w.r.t. $d$ and is computable in $O(h(n) \cdot n^{1+1/h(n)})$ time. However, $\tilde{d}(\cdot, \cdot)$ depends on $\Theta(h(n) \cdot n^{1+1/h(n)})$ distances of $\tilde{d}$, forbidding us to improve the query complexity of $O(h(n) \cdot n^{1+1/h(n)})$ in Theorem 1. Wu’s [12] algorithm also makes $\Theta(h(n) \cdot n^{1+1/h(n)})$ queries. In contrast, our main contribution is a new function, $\hat{d}$, that depends on only $\Theta(n^{1+1/h(n)})$ distances of $\tilde{d}$ and is otherwise similar to Chang’s $\tilde{d}$. This results in a deterministic, $O(h(n) \cdot n^{1+1/h(n)})$-time, $O(n^{1+1/h(n)})$-query, $(2h(n))$-approximation and nonadaptive algorithm for METRIC 1-MEDIAN, improving the query complexity in Theorem 1. The idea behind our construction of $\hat{d}$ comes from an unpublished workshop paper of Chang [3]. Aside from our design of $\hat{d}$, most of our derivations are simple modifications of those of Chang [2]. As a corollary to our result, METRIC 1-MEDIAN has a deterministic, $O(n \log n)$-time, $O(n)$-query, $(\epsilon \log n)$-approximation and nonadaptive algorithm for all constants $\epsilon > 0$.

On the negative side, Chang [4, 5] proves that METRIC 1-MEDIAN has no deterministic $o(n^{1+1/(h(n)-1)/h(n)})$-query $(2h(n) - \epsilon)$-approximation algorithms for any $\epsilon > 0$ and any $h: \mathbb{Z}^+ \to \mathbb{Z}^+-\{1\}$ satisfying $h(n) = o(n^{1/(h(n)-1)})$. So there is still gap between Chang’s lower bound and our upper bound of $O(n^{1+1/h(n)})$ on the query complexity.

### 2 Our pseudo-distance function

Let $(\{0, 1, \ldots, n-1\}, d)$ be a metric space and $h: \mathbb{Z}^+ \to \mathbb{Z}^+-\{1\}$ be a computable function. By Bertrand’s postulate, there exists a prime number $t \in [\lceil n^{1/h(n)} \rceil, 2 \cdot \lceil n^{1/h(n)} \rceil]$. Clearly, $\gcd(n-1, n) = 1$. So the primality of $t$ implies the existence of $\sigma \in \{0, 1\}$ such that $\gcd(t, n-\sigma) = 1$. For convenience, $h \equiv h(n)$. For all $j \in \{0, 1, \ldots, n-1\}$, write

$$(s_{h-1}(j), s_{h-2}(j), \ldots, s_0(j)) \in \{0, 1, \ldots, t-1\}^h$$

for the unique $t$-ary representation of $j$, following Chang [2]. So

$$\sum_{\ell=0}^{h-1} s_{h-1-\ell}(j) \cdot t^{h-1-\ell} = j. \quad (1)$$

For any predicate $P$, let $\chi[P] = 1$ if $P$ is true and $\chi[P] = 0$ otherwise.
Define
\[ d^{(n-\sigma)}(x, y) \equiv d(x \mod (n - \sigma), y \mod (n - \sigma)) \] (2)
for all \( x, y \in \mathbb{N} \). Clearly, \( d^{(n-\sigma)} \) is symmetric and obeys the triangle inequality, just like \( d \). For all \( i, j \in \{0, 1, \ldots, n - \sigma - 1\} \), define
\[
\hat{d}(i, it^h + j \mod (n - \sigma)) \\
= \sum_{k=0}^{h-1} d^{(n-\sigma)}(it^k + \sum_{\ell=0}^{k-1} s_{h-1-\ell}(j) \cdot t^{k-1-\ell}, it^{k+1} + \sum_{\ell=0}^{k} s_{h-1-\ell}(j) \cdot t^{k-\ell}) .
\] (3)
This and the triangle inequality for \( d^{(n-\sigma)} \) imply
\[
\hat{d}(i, it^h + j \mod (n - \sigma)) \geq d^{(n-\sigma)}(i, it^h + j) .
\] (4)

So by equation (1),
\[
\hat{d}(i, it^h + j \mod (n - \sigma)) \geq d^{(n-\sigma)}(i, it^h + j) .
\] (4)

Note that the domain of \( \hat{d} \) is \( \{0, 1, \ldots, n - \sigma - 1\}^2 \). For all \( x, y \in \mathbb{N} \), interpret \( \hat{d}(x, y) \) as
\[
\hat{d}(x \mod (n - \sigma), y \mod (n - \sigma)) .
\]

Let
\[
i' = \arg\min_{i=0}^{n-\sigma-1} \sum_{j=0}^{n-1} d(i, j),
\] (5)
breaking ties arbitrarily.

When \( \sigma = 0 \), the following lemma says that a 1-median w.r.t. \( \hat{d} \) is a \((2h)\)-approximate 1-median w.r.t. \( d \).

**Lemma 2** (cf. [2, Lemma 4]). Let
\[
\alpha = \arg\min_{i=0}^{n-\sigma-1} \left( \chi[\sigma = 1] \cdot d(i, n - 1) + \sum_{j=0}^{n-\sigma-1} d(i, it^h + j \mod (n - \sigma)) \right),
\] (6)
breaking ties arbitrarily. Then
\[
\sum_{j=0}^{n-1} d(\alpha, j) \\
\leq \chi[\sigma = 1] \cdot d(\alpha, n - 1) + \sum_{j=0}^{n-\sigma-1} \hat{d}(\alpha, it^h + j \mod (n - \sigma)) \\
\leq 2h \cdot \left( \min_{i=0}^{n-\sigma-1} \sum_{j=0}^{n-1} d(i, j) \right) - \chi[\sigma = 1] \cdot \left( (2h - 1) \cdot d(i', n - 1) - \frac{1}{n-1} \sum_{j=0}^{n-2} d(i', j) \right).
\]
Proof. Clearly,
\[
\sum_{j=0}^{n-1} d(\alpha,j) \overset{\text{(2)}}{=} \chi[\sigma=1] \cdot d(\alpha,n-1) + \sum_{j=0}^{n-\sigma-1} d^{(n-\sigma)}(\alpha,j)
= \chi[\sigma=1] \cdot d(\alpha,n-1) + \sum_{j=0}^{n-\sigma-1} d^{(n-\sigma)}(\alpha,\alpha t^h + j),
\]
where the second equality uses equation (2) and the one-to-one correspondence of \( j \mapsto \alpha t^h + j \mod (n-\sigma) \) for \( j \in \{0,1,\ldots,n-\sigma-1\} \).

Pick \( u \) from \( \{0,1,\ldots,n-\sigma-1\} \) uniformly at random. Then
\[
\chi[\sigma=1] \cdot d(\alpha,n-1) + \sum_{j=0}^{n-\sigma-1} d(\alpha,\alpha t^h + j) \leq \chi[\sigma=1] \cdot d(u,n-1) + \sum_{j=0}^{n-\sigma-1} d(u,ut^h + j \mod (n-\sigma)) \leq \chi[\sigma=1] \cdot (d(i',u) + d(i',n-1)) + \sum_{j=0}^{n-\sigma-1} d(u,ut^h + j \mod (n-\sigma)) = \chi[\sigma=1] \cdot \left( \frac{1}{n-\sigma} \cdot \sum_{m=0}^{n-\sigma-1} d(i',m) + d(i',n-1) \right) + \mathbb{E} \left[ \sum_{j=0}^{n-\sigma-1} d(u,ut^h + j \mod (n-\sigma)) \right],
\]
where the last inequality follows from the triangle inequality for \( d \). Furthermore,
\[
\mathbb{E} \left[ \sum_{j=0}^{n-\sigma-1} d(u,ut^h + j \mod (n-\sigma)) \right] \overset{\text{(3)}}{=} \mathbb{E} \left[ \sum_{j=0}^{n-\sigma-1} \sum_{k=0}^{h-1} d^{(n-\sigma)}(ut^k + \sum_{\ell=0}^{k-1} s_{h-1-\ell}(j) \cdot t^{k-1-\ell} \cdot t^k+1 + \sum_{\ell=0}^{k-1} s_{h-1-\ell}(j) \cdot t^k) \right] \leq \mathbb{E} \left[ \sum_{j=0}^{h-1} \sum_{k=0}^{h-1} d^{(n-\sigma)}(i',ut^k + \sum_{\ell=0}^{k-1} s_{h-1-\ell}(j) \cdot t^{k-1-\ell} \cdot t^k) + d^{(n-\sigma)}(i',ut^{k+1} + \sum_{\ell=0}^{k} s_{h-1-\ell}(j) \cdot t^{k-\ell}) \right] \overset{\text{(2)}}{=} \sum_{j=0}^{n-\sigma-1} \sum_{k=0}^{h-1} \left( \mathbb{E} \left[ d(i',ut^k + \sum_{\ell=0}^{k-1} s_{h-1-\ell}(j) \cdot t^{k-1-\ell} \mod (n-\sigma)) \right] \right) + \mathbb{E} \left[ d(i',ut^{k+1} + \sum_{\ell=0}^{k} s_{h-1-\ell}(j) \cdot t^{k-\ell} \mod (n-\sigma)) \right]
\]
where the inequality follows from the triangle inequality for \( d^{(n-\sigma)} \).

Because \( u \) is a uniformly random element of \( \{0,1,\ldots,n-\sigma-1\} \) and \( \gcd(t,n-
Therefore, for any $j \in \{0, 1, \ldots, n-1\}$ and $k \in \{0, 1, \ldots, h\}$.

\[
\sum_{j=0}^{n-\sigma-1} \sum_{k=0}^{h-1} \left( \frac{1}{n-\sigma} \cdot \sum_{m=0}^{n-\sigma-1} d(i', m) + \frac{1}{n-\sigma} \cdot \sum_{m=0}^{n-\sigma-1} d(i', m) \right)
= 2h \sum_{m=0}^{n-\sigma-1} d(i', m).
\]

Summarizing all the above with tedious calculations,

\[
\sum_{j=0}^{n-1} d(\alpha, j) \leq \chi[\sigma = 1] \cdot d(\alpha, n-1) + \sum_{j=0}^{n-\sigma-1} \hat{d}(\alpha, \alpha^h + j \mod (n-\sigma)) \\
\leq 2h \left( \sum_{m=0}^{n-1} d(i', m) \right) - \chi[\sigma = 1] \cdot \left( (2h-1) \cdot d(i', n-1) - \frac{1}{n-1} \cdot \sum_{m=0}^{n-2} d(i', m) \right).
\]

Finally, invoke equation (5).

The following lemma shows how to pick a $(2h)$-approximate $1$-median (w.r.t. $d$) from $\{\alpha, n-1\}$.

**Lemma 3.** Let $\alpha \in \{0, 1, \ldots, n - \sigma - 1\}$ be as in equation (6), breaking ties arbitrarily. If

\[
\chi[\sigma = 1] \cdot d(\alpha, n-1) + \sum_{j=0}^{n-\sigma-1} \hat{d}(\alpha, \alpha^h + j \mod (n-\sigma)) < \sum_{j=0}^{n-1} d(n-1, j),
\]

then

\[
\sum_{j=0}^{n-1} d(\alpha, j) \leq 2h \cdot \min_{i=0}^{n-1} \sum_{j=0}^{n-1} d(i, j).
\]

Otherwise,

\[
\sum_{j=0}^{n-1} d(n-1, j) \leq 2h \cdot \min_{i=0}^{n-1} \sum_{j=0}^{n-1} d(i, j).
\]
Proof. Clearly,

\[ \sum_{j=0}^{n-1} d(\alpha, j) = \chi[\sigma = 1] \cdot d(\alpha, n-1) + \sum_{j=0}^{n-\sigma-1} d^{(n-\sigma)}(\alpha, j) \] (10)

\[ \leq \chi[\sigma = 1] \cdot d(\alpha, n-1) + \sum_{j=0}^{n-\sigma-1} \hat{d}(\alpha, j) \]

\[ = \chi[\sigma = 1] \cdot d(\alpha, n-1) + \sum_{j=0}^{n-\sigma-1} \hat{d}(\alpha, \alpha h + j \mod (n-\sigma)), \] (11)

where the last equality uses the one-to-one correspondence of \( j \mapsto \alpha h + j \mod (n-\sigma) \) for \( j \in \{0, 1, \ldots, n-\sigma-1\} \).

Next, we separate the discussion as to whether

\[ (2h-1) \cdot d(\alpha', n-1) - \frac{1}{n-1} \sum_{m=0}^{n-2} d(\alpha', m) \geq 0. \] (12)

Case (1): Equation (12) is true. By Lemma 2,

\[ \sum_{j=0}^{n-1} d(\alpha, j) \leq 2h \cdot \min_{i=0}^{n-\sigma-1} \sum_{j=0}^{n-1} d(i, j). \] (13)

Subcase (i): Equation (7) is true. By equations (7) and (10)–(11),

\[ \sum_{j=0}^{n-1} d(\alpha, j) < \sum_{j=0}^{n-1} d(n-1, j). \]

This and equation (13) imply equation (8).

Subcase (ii): Equation (7) is false. By Lemma 2 and equation (12),

\[ \chi[\sigma = 1] \cdot d(\alpha, n-1) + \sum_{j=0}^{n-\sigma-1} \hat{d}(\alpha, \alpha h + j \mod (n-\sigma)) \]

\[ \leq 2h \cdot \min_{i=0}^{n-\sigma-1} \sum_{j=0}^{n-1} d(i, j). \]

This and the negation of equation (7) imply

\[ \sum_{j=0}^{n-1} d(n-1, j) \leq 2h \cdot \min_{i=0}^{n-\sigma-1} \sum_{j=0}^{n-1} d(i, j). \] (14)

Equation (14) implies equation (9) (note that \( \sum_{j=0}^{n-1} d(n-1, j) \) does not exceed itself).
Case (2): Equation (12) is false. By the triangle inequality for $d$,
\[
\sum_{j=0}^{n-1} d(n-1,j) \leq \sum_{j=0}^{n-1} (d(i',n-1) + d(i',j)) = n \cdot d(i',n-1) + \sum_{j=0}^{n-1} d(i',j).
\]
This and the negation of equation (12) imply
\[
\sum_{j=0}^{n-1} d(n-1,j) < n \cdot \frac{1}{2h-1} \cdot \frac{1}{n-1} \sum_{j=0}^{n-2} d(i',j) + \sum_{j=0}^{n-1} d(i',j) \tag{15}
\]
\[
\leq 2 \cdot \sum_{j=0}^{n-1} d(i',j) \tag{16}
\]
where the second inequality uses $h \geq 2$. Inequalities (15)–(16) imply equation (9).

Subcase (a): Equation (7) is false. Equation (9) holds as desired.
Subcase (b): Equation (7) is true. By equations (7) and (10)–(11),
\[
\sum_{j=0}^{n-1} d(\alpha,j) < \sum_{j=0}^{n-1} d(n-1,j).
\]
This and equation (9) give equation (8).

3 Dynamic programming

Define $(s'_h, s'_{h-1}, \ldots, s'_0) \in \{0, 1, \ldots, t-1\}^h$ to be the $t$-ary representation of $n - \sigma - 1$. So $\sum_{r=0}^{h-1} s'_r \cdot t^r = n - \sigma - 1$. For $i \in \{0, 1, \ldots, n - \sigma - 1\}$ and $m \in \{0, 1, \ldots, h-1\}$, define
\[
f(i,m) \equiv \sum_{s_m, s_{m-1}, \ldots s_0 = 0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t^r \right] 
\cdot \sum_{k=0}^{m} d^{(n-\sigma)} \left( it^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell} \cdot it^k + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right), \tag{17}
\]
\[
g(i,m) \equiv \sum_{s_m, s_{m-1}, \ldots s_0 = 0}^{t-1} \sum_{k=0}^{m} d^{(n-\sigma)} \left( it^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell} \cdot it^k + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right); \tag{18}
\]
hence

\[ f(i, 0) = \sum_{s_0=0}^{s'_0} d^{(n-\sigma)}(i, it + s_0), \quad (19) \]

\[ g(i, 0) = \sum_{s_0=0}^{t-1} d^{(n-\sigma)}(i, it + s_0). \quad (20) \]

Chang also defines functions similar to our \( f(\cdot, \cdot) \) and \( g(\cdot, \cdot) \) [2, equations (8)–(9)], based on his pseudo-distance function [2, equation (2)]. Instead, equations (17)–(18) are based on \( \hat{d} \) in equation (3).

When \( \sigma = 0 \), the following lemma says that a minimizer of \( f(\cdot, h - 1) \) is a 1-median w.r.t. \( \hat{d} \).

**Lemma 4** (cf. [2, Lemma 5]). For all \( i \in \{0, 1, \ldots, n - \sigma - 1\} \),

\[ f(i, h - 1) = \sum_{j=0}^{n-\sigma-1} \hat{d}(i, it^h + j \mod (n - \sigma)). \]

**Proof.** Representing each \( \sum_{j=0}^{n-\sigma-1} \hat{d}(i, it^h + j \mod (n - \sigma)) \) in \( t \)-ary as \( (s_{h-1}, s_{h-2}, \ldots, s_0) \),

\[
\begin{align*}
\sum_{j=0}^{n-\sigma-1} \sum_{k=0}^{h-1} d^{(n-\sigma)} \left( it^k + \sum_{\ell=0}^{k-1} s_{h-1-\ell}(j) \cdot t^{k-1-\ell}, it^{k+1} + \sum_{\ell=0}^{k} s_{h-1-\ell}(j) \cdot t^{k-\ell} \right) \\
= \sum_{s_{h-1},s_{h-2},\ldots,s_0=0}^{h-1} \chi \left( \sum_{r=0}^{k-1} s_r \cdot t^r \leq n - \sigma - 1 \right) \\
\cdot \sum_{k=0}^{h-1} d^{(n-\sigma)} \left( it^k + \sum_{\ell=0}^{k-1} s_{h-1-\ell} \cdot t^{k-1-\ell}, it^{k+1} + \sum_{\ell=0}^{k} s_{h-1-\ell} \cdot t^{k-\ell} \right). \quad (21)
\end{align*}
\]

Equations (3), (17) and (21) complete the proof (recall that \( \sum_{r=0}^{h-1} s'_r \cdot t^r = n - \sigma - 1 \)).

When \( \sigma = 0 \), a minimizer of \( f(\cdot, h - 1) \) is a \( (2h) \)-approximate 1-median w.r.t. \( \hat{d} \) by Lemmas 2 and 4. So we want to calculate \( f(\cdot, h - 1) \). The next four lemmas derive recurrences for \( g(\cdot, \cdot) \) and \( f(\cdot, \cdot) \).

**Lemma 5** (cf. [2, Lemma 6]). For all \( m \in \{0, 1, \ldots, h - 1\} \) and \( s_m, s_{m-1}, \ldots, s_0 \in \{0, 1, \ldots, t - 1\} \),

\[
\begin{align*}
\sum_{k=1}^{m} d^{(n-\sigma)} \left( it^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, it^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \\
= \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( it^{k+1} + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-1-\ell} \cdot t^{k-1-\ell}, it^{k+2} + s_m \cdot t^{k+1} + \sum_{\ell=0}^{k} s_{m-1-\ell} \cdot t^{k-\ell} \right)
\end{align*}
\]
Proof. Clearly,
\[\sum_{k=1}^{m} d^{(n-s)} \left( ik^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, ik^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \]
\[= \sum_{k=0}^{m-1} d^{(n-s)} \left( ik^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell}, ik^{k+2} + \sum_{\ell=0}^{k+1} s_{m-\ell} \cdot t^{k+1-\ell} \right) \]
\[= \sum_{k=0}^{m-1} d^{(n-s)} \left( ik^{k+1} + s_{m} \cdot t^{k} + \sum_{\ell=1}^{k-1} s_{m-\ell} \cdot t^{k-\ell}, ik^{k+2} + s_{m} \cdot t^{k+1} + \sum_{\ell=1}^{k+1} s_{m-\ell} \cdot t^{k+1-\ell} \right) \]
\[= \sum_{k=0}^{m-1} d^{(n-s)} \left( ik^{k+1} + s_{m} \cdot t^{k} + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, ik^{k+2} + s_{m} \cdot t^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right), \]
where the first and the last equalities follow from substituting \(k\) with \(k+1\) and \(\ell\) with \(\ell+1\), respectively. \(\square\)

Lemma 6 (cf. [2, Lemma 6]). For all \(i \in \{0, 1, \ldots, n - \sigma - 1\}\) and \(m \in \{1, 2, \ldots, h - 1\}\),
\[g(i, m) = m \sum_{s_{m}=0}^{t-1} d^{(n-s)} (i, it + s_{m}) \]
\[+ \sum_{s_{m}=0}^{t-1} g(it + s_{m} \mod (n - \sigma), m - 1). \]

Proof. By equation (18),
\[g(i, m) \]
\[= \sum_{s_{m}=0}^{t-1} \sum_{s_{m-1}, s_{m-2}, \ldots, s_{0}=0} d^{(n-s)} (i, it + s_{m}) \]
\[+ \sum_{k=1}^{m} d^{(n-s)} \left( ik^{k} + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, ik^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \]
\[= \sum_{s_{m}=0}^{t-1} t^{m} \cdot d^{(n-s)} (i, it + s_{m}) \]
\[+ \sum_{s_{m}=0}^{t-1} \sum_{s_{m-1}, s_{m-2}, \ldots, s_{k}=0}^{t-1} \sum_{k=1}^{m} d^{(n-s)} \left( ik^{k} + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, ik^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right), \]
\[g(it + s_{m} \mod (n - \sigma), m - 1) \]
\[= \sum_{s_{m}=0}^{t-1} \sum_{s_{m-1}, s_{m-2}, \ldots, s_{0}=0}^{m-1} d^{(n-s)} \left( it + s_{m} \cdot t^{k} + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, (it + s_{m}) \cdot t^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \]
for \(s_{m} \in \{0, 1, \ldots, t-1\}\), where the last equality uses equation (2) as well. These and Lemma 5 complete the proof. \(\square\)
For all \( m \in \{1, 2, \ldots, h - 1\} \) and \( s_m, s_{m-1}, \ldots, s_0 \in \{0, 1, \ldots, t-1\} \), consider whether

\[
\sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t^r
\]

(22)

in the following three cases:

(I) If \( s_m = s'_m \), equation (22) holds if and only if \((s_{m-1}, s_{m-2}, \ldots, s_0)\) is the \(t\)-ary representation of one of \(0, 1, \ldots, \sum_{r=0}^{m-1} s'_r \cdot t^r\).

(II) If \( s_m < s'_m \), equation (22) holds.

(III) If \( s_m > s'_m \), equation (22) fails to hold.

**Lemma 7** (cf. [2, Lemma 7]). For all \( m \in \{1, 2, \ldots, h - 1\} \),

\[
\sum_{s_m=0}^{s'_m} \sum_{s_{m-1}, s_{m-2}, \ldots, s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t^r \right] \cdot d^{(n-\sigma)}(i, it + s_m)
\]

\[
= \left(1 + \sum_{r=0}^{m-1} s'_r \cdot t^r\right) d^{(n-\sigma)}(i, it + s'_m) + t^m \sum_{s_m=0}^{s'_m-1} d^{(n-\sigma)}(i, it + s_m).
\]

(23)

**Proof.** Items (I)–(II) account for the first and the second terms of the right-hand side of equation (23), respectively. \(\square\)

**Lemma 8** (cf. [2, Lemma 7]). For all \( i \in \{0, 1, \ldots, n - \sigma - 1\} \) and \( m \in \{1, 2, \ldots, h - 1\} \),

\[
f(i, m) = \left(1 + \sum_{r=0}^{m-1} s'_r \cdot t^r\right) d^{(n-\sigma)}(i, it + s'_m)
\]

\[
+ t^m \sum_{s_m=0}^{s'_m-1} d^{(n-\sigma)}(i, it + s_m)
\]

\[
+ f(it + s'_m \mod (n - \sigma), m - 1)
\]

\[
+ \sum_{s_m=0}^{s'_m-1} g(it + s_m \mod (n - \sigma), m - 1).
\]
Proof. We have

\begin{align}
  f(i, m) \\
  \overset{(17)}{=} \\
  \sum_{s_m=0}^{t-1} \sum_{s_{m-1},s_{m-2},...,s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t' \right] \cdot \left( d^{(n-\sigma)} (i, it + s_m) \right) \\
  + \sum_{k=1}^{m} d^{(n-\sigma)} \left( it^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, it^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \\
  \overset{Lemma \ 5}{=} \\
  \sum_{s_m=0}^{m-1} \sum_{s_{m-1},s_{m-2},...,s_0=0}^{m-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t' \right] \cdot \left( d^{(n-\sigma)} (i, it + s_m) \right) \\
  + \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( it^{k+1} + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, it^{k+2} + s_m \cdot t^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \\
  \overset{item \ (III)}{=} \\
  \sum_{s_m=0}^{s'_m} \sum_{s_{m-1},s_{m-2},...,s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t' \right] \cdot \left( d^{(n-\sigma)} (i, it + s_m) \right) \\
  + \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( it^{k+1} + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, it^{k+2} + s_m \cdot t^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \\
  \overset{Lemma \ 7}{=} \\
  \sum_{s_m=0}^{s'_m} \sum_{s_{m-1},s_{m-2},...,s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t' \right] \cdot d^{(n-\sigma)} \left( it + s'_m \mod (n-\sigma), m-1 \right) + \sum_{s_m=0}^{s'_m-1} g \left( it + s_m \mod (n-\sigma), m-1 \right).
\end{align}

So by Lemma 7, it remains to prove that

\begin{align}
  \sum_{s_m=0}^{s'_m} \sum_{s_{m-1},s_{m-2},...,s_0=0}^{m-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_r \cdot t' \right] \\
  \cdot \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( it^{k+1} + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-1-\ell}, it^{k+2} + s_m \cdot t^{k+1} + \sum_{\ell=0}^{k} s_{m-\ell} \cdot t^{k-\ell} \right) \\
  = f \left( it + s'_m \mod (n-\sigma), m-1 \right) + \sum_{s_m=0}^{s'_m-1} g \left( it + s_m \mod (n-\sigma), m-1 \right).
\end{align}

Separating the left-hand side of equation (25) according to whether \( s_m = s'_m \).
or \( s_m \leq s'_{m-1} - 1, \)

\[
\sum_{s_m=0}^{s'_{m-1}} \sum_{s_{m-1}=0, s_{m-2}, \ldots, s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_{r} \cdot t^r \right] 
\cdot \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( (it+1) + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-\ell}, \right. \\
\left. (it+1) + s'_{m-1} \cdot t^{k-1}, \ldots, \right. \\
\left. (it+1) + s'_{m} \cdot t^{k-\ell}, \right.
\left. (it+1) + s'_{m-1-\ell} \cdot t^{k-\ell} \right)
\]

\[
= \sum_{s_{m-1}, s_{m-2}, \ldots, s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_{r} \cdot t^r \right] 
\cdot \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( (it+1) + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-\ell}, \right. \\
\left. (it+1) + s'_{m-1} \cdot t^{k-1}, \ldots, \right. \\
\left. (it+1) + s'_{m} \cdot t^{k-\ell}, \right.
\left. (it+1) + s'_{m-1-\ell} \cdot t^{k-\ell} \right)
\]

\[
\cdot \sum_{s_{m}=0}^{s'_{m-1}} \sum_{s_{m-1}, s_{m-2}, \ldots, s_0=0}^{t-1} \chi \left[ \sum_{r=0}^{m} s_r \cdot t^r \leq \sum_{r=0}^{m} s'_{r} \cdot t^r \right] 
\cdot \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( (it+1) + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-\ell}, \right. \\
\left. (it+1) + s'_{m-1} \cdot t^{k-1}, \ldots, \right. \\
\left. (it+1) + s'_{m} \cdot t^{k-\ell}, \right.
\left. (it+1) + s'_{m-1-\ell} \cdot t^{k-\ell} \right)
\]

\[
\cdot \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( (it+1) + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-\ell}, \right. \\
\left. (it+1) + s'_{m-1} \cdot t^{k-1}, \ldots, \right. \\
\left. (it+1) + s'_{m} \cdot t^{k-\ell}, \right.
\left. (it+1) + s'_{m-1-\ell} \cdot t^{k-\ell} \right)
\]

\[
\cdot \sum_{k=0}^{m-1} d^{(n-\sigma)} \left( (it+1) + s_m \cdot t^k + \sum_{\ell=0}^{k-1} s_{m-\ell} \cdot t^{k-\ell}, \right. \\
\left. (it+1) + s'_{m-1} \cdot t^{k-1}, \ldots, \right. \\
\left. (it+1) + s'_{m} \cdot t^{k-\ell}, \right.
\left. (it+1) + s'_{m-1-\ell} \cdot t^{k-\ell} \right)
\]

\[
(17)-(18) \quad f \left( it + s'_{m} \mod (n-\sigma), m-1 \right) + \sum_{s_m=0}^{s'_{m-1}} g \left( it + s_m \mod (n-\sigma), m-1 \right). \]
Lemma 9 (cf. [2,Lemma 8]). Approx.-Median in Fig. 1 outputs a (2h)-approximate 1-median w.r.t. d.

Proof. By equations (19)–(20), lines 10–13 of Approx.-Median find $f(\cdot, 0)$ and $g(\cdot, 0)$. By Lemmas 6 and 8, lines 14–23 find $f(\cdot, m)$ and $g(\cdot, m)$ for an increasing $m \in \{1, 2, \ldots, h - 1\}$. By Lemma 4, line 24 picks $\alpha$ as in equation (6), and the condition in line 25 is the same as equation (7). So by Lemma 3, lines 25–29 output a (2h)-approximate 1-median (w.r.t. d).

Below is our main theorem.

Theorem 10 (cf. [2,Theorem 9]). Let $h: \mathbb{Z}^+ \rightarrow \mathbb{Z}^+ \setminus \{1\}$ be such that (1) $h(n) \leq \lg n$ for all sufficiently large $n$ and (2) $h(n)$ and $\lceil n^{1/h(n)} \rceil$ are computable from $n$ in $O(h(n) \cdot n^{1+1/h(n)})$ time. Then METRIC 1-MEDIAN has a deterministic, $O(h(n) \cdot n^{1+1/h(n)})$-time, $O(n^{1+1/h(n)})$-query, (2h(n))-approximation and nonadaptive algorithm.

Proof. By Lemma 9, Approx.-Median is (2h)-approximate. It is clearly deterministic and nonadaptive. Lines 3–8 of Approx.-Median make $O(nt)$ queries. With $\{t'_i\}_{i=0}^{h-1}$ and $\{\sum_{r=0}^{i} s'_r \cdot t'_r\}_{i=0}^{h-1}$ precomputed, lines 14–23 take $O(hnt)$ time. The well-known AKS primality test allows line 1 to take time

$$\left(\lceil n^{1/h} \rceil + 1\right) \cdot \log^{O(1)} \left( O \left( n^{1/h} \right) \right) = O \left( \sqrt{n} \cdot \log^{O(1)} n \right) = o(n),$$

where the first equality uses $h \geq 2$. Finally, $t = \Theta(n^{1/h})$ by line 1.

Corollary 11. METRIC 1-MEDIAN has a deterministic, $O(n \log n)$-time, $O(n)$-query, $(\epsilon \log n)$-approximation and nonadaptive algorithm for each constant $\epsilon > 0$.

Proof. Take $h(n) = (\epsilon/2) \lg n$ in Theorem 10.

Corollary 11 is stronger than taking $h(n) = (\epsilon/2) \lg n$ in Theorem 1. We leave open whether deterministic $O(n)$-query algorithms for METRIC 1-MEDIAN can be $o(\log n)$-approximate.

Acknowledgments

The author is supported in part by the Ministry of Science and Technology of Taiwan under grant 105-2221-E-155-047-.
1: Pick any prime number \( t \in \lceil n^{1/h} \rceil, 2 \cdot \lceil n^{1/h} \rceil \);

2: Pick any \( \sigma \in \{0, 1\} \) satisfying \( \gcd(t, n - \sigma) = 1 \);

3: for \( i = 0, 1, \ldots, n - \sigma - 1 \) do

4: for \( s = 0, 1, \ldots, t - 1 \) do

5: Query for \( d(i, it + s \mod (n - \sigma)) \);

6: end for

7: Query for \( d(n - 1, i) \);

8: end for

9: \((s'_{h-1}, s'_{h-2}, \ldots, s'_0) \) ← the \( t \)-ary representation of \( n - \sigma - 1 \);

10: for \( i = 0, 1, \ldots, n - \sigma - 1 \) do

11: \( f[i][0] \leftarrow \sum_{s_0=0}^{s'_0} d(i, it + s_0 \mod (n - \sigma)) \);

12: \( g[i][0] \leftarrow \sum_{s_0=0}^{t-1} d(i, it + s_0 \mod (n - \sigma)) \);

13: end for

14: for \( m = 1, 2, \ldots, h - 1 \) do

15: for \( i = 0, 1, \ldots, n - \sigma - 1 \) do

16: \( f[i][m] \leftarrow (1 + \sum_{r=0}^{m-1} s'_r \cdot t^r) d(i, it + s'_m \mod (n - \sigma)) \);

17: \( f[i][m] \leftarrow f[i][m] + t^m \sum_{s_m=0}^{s'_m-1} d(i, it + s_m \mod (n - \sigma))[m-1] \);

18: \( f[i][m] \leftarrow f[i][m] + f[it + s'_m \mod (n - \sigma)][m-1] \);

19: \( f[i][m] \leftarrow f[i][m] + \sum_{s_m=0}^{s'_m-1} g[it + s_m \mod (n - \sigma)][m-1] \);

20: \( g[i][m] \leftarrow t^m \sum_{s_m=0}^{t-1} d(i, it + s_m \mod (n - \sigma)) \);

21: \( g[i][m] \leftarrow g[i][m] + \sum_{s_m=0}^{t-1} g[it + s_m \mod (n - \sigma)][m-1] \);

22: end for

23: end for

24: \( \alpha \leftarrow \arg\min_{i=0}^{n-\sigma-1} (\chi[\sigma = 1] \cdot d(i, n - 1) + f[i][h - 1]) \), breaking ties arbitrarily;

25: if \( \chi[\sigma = 1] \cdot d(\alpha, n - 1) + f[\alpha][h - 1] < \sum_{j=0}^{n-1} d(n - 1, j) \) then

26: Output \( \alpha \);

27: else

28: Output \( n - 1 \);

29: end if

Figure 1: Algorithm Approx.-Median
References

[1] C.-L. Chang. Deterministic sublinear-time approximations for metric 1-median selection. *Information Processing Letters*, 113(8):288–292, 2013.

[2] C.-L. Chang. A deterministic sublinear-time nonadaptive algorithm for metric 1-median selection. *Theoretical Computer Science*, 602:149–157, 2015.

[3] C.-L. Chang. A note on 1-median selection in metric spaces. In *Proceedings of the 32nd Workshop on Combinatorial Mathematics and Computation Theory*, pages 51–54, Taichung, Taiwan, 2015. [http://par.cse.nsus.edu.tw/~algo/paper/paper15/B1_3.pdf](http://par.cse.nsus.edu.tw/~algo/paper/paper15/B1_3.pdf) (This workshop does not publish proceedings.).

[4] C.-L. Chang. Metric 1-median selection: Query complexity vs. approximation ratio. In *Proceedings of the 22nd International Computing and Combinatorics Conference*, pages 131–142, Ho Chi Minh City, Vietnam, 2016. [https://arxiv.org/abs/1509.05662](https://arxiv.org/abs/1509.05662).

[5] C.-L. Chang. A lower bound for metric 1-median selection. *Journal of Computer and System Sciences*, 84:44–51, 2017.

[6] K. Chen. On coresets for $k$-median and $k$-means clustering in metric and Euclidean spaces and their applications. *SIAM Journal on Computing*, 39(3):923–947, 2009.

[7] S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O’Callaghan. Clustering data streams: Theory and practice. *IEEE Transactions on Knowledge and Data Engineering*, 15(3):515–528, 2003.

[8] P. Indyk. Sublinear time algorithms for metric space problems. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 428–434, 1999.

[9] P. Indyk. *High-dimensional computational geometry*. PhD thesis, Stanford University, 2000.

[10] A. Kumar, Y. Sabharwal, and S. Sen. Linear-time approximation schemes for clustering problems in any dimensions. *Journal of the ACM*, 57(2):5, 2010.

[11] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 3rd edition, 1976.

[12] B.-Y. Wu. On approximating metric 1-median in sublinear time. *Information Processing Letters*, 114(4):163–166, 2014.