ELSEVIER

Information Processing Letters 74 (2000) 103–106

Information Processing Letters

www.elsevier.com/locate/ipl

# On finding medians of weighted discrete points

Kuo-Liang Chung [1]

*Department of Information Management, Institute of Computer Science and Information Engineering, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei, Taiwan 10672*

Received 13 January 2000; received in revised form 2 March 2000
Communicated by L. Boasson

**Abstract**

According to the Manhattan metric, Del Lungo et al. (1998) recently presented an elegant algorithm for finding the medians of the given discrete point set $S$ on $\mathbb{Z}^2$, where each point is of unit weight. Their algorithm takes $O(|R(S)|)$ time, where $|R(S)|$ denotes the area of the smallest rectangle containing $S$. Under the same time bound, this paper first extends their result to the weighted case, where each point in $S$ is associated with any weight. Secondly, we use the sparse matrix representation technique to find the medians of $S$ in $O(|S|)$ time, i.e., linear time, commonly $|S| \leqslant |R(S)|$. 

*Keywords:* Algorithms; Computational geometry; Manhattan metric; Medians; Weighted discrete set

## 1. Introduction

Given a set of discrete points on the 2D integer domain $\mathbb{Z}^2$, say $S$, each point with unit weight, finding the medians of $S$ is an interesting problem in computational geometry [2,3]. In some applications, for instance, the found medians could be used as the central servers for those $|S|$ discrete nodes in the environment. Using the Manhattan metric, the distance between $S$ and one position $p = (x_p, y_p)$ on $\mathbb{Z}^2$ is defined by

$$D(p, S) = \sum_{q \in S} d_1(p, q),$$

where

$$d_1(p, q) = |x_p - x_q| + |y_p - y_q|.$$

A position $m$ on $\mathbb{Z}^2$ is said to be a median of $S$ if the equality

$$D(m, S) = \min_{q \in \mathbb{Z}^2} D(q, S)$$

holds. Therefore, usually more than one median will be found and some medians on $\mathbb{Z}^2$ do not belong to $S$.

Recently, Del Lungo et al. [4] presented a very elegant method for finding the medians of $S$. Their algorithm takes $O(|R(S)|)$ time, where $|R(S)|$ denotes the area of the smallest rectangle containing $S$, commonly $|S| \leqslant |R(S)|$. From a practical viewpoint, we often consider a set of weighted discrete points, still say $S$. The motivations of this research are twofold: extending the result of Del Lungo et al. [4] to the weighted case and presenting a linear-time (i.e., $O(|S|)$ time) algorithm for finding the medians of $S$.

Using the same Manhattan metric, this paper first extends the result of Del Lungo et al. [4] to the weighted case, where the discrete points in $S$ are

[1] Email: klchung@cs.ntust.edu.tw. Supported by NSC89-2213-E011-061.

0020-0190/00/$ – see front matter 
PII: S0020-0190(00)00049-1

weighted. Furthermore, using the sparse matrix representation technique, the problem of finding medians can be solved in $\mathrm{O}(|S|)$ time.

## 2. Weighted discrete points

Let the discrete point $q$ in $S$ have weight $w_q \in \mathbb{Z}$, the distance between the position $p$ on $\mathbb{Z}^2$ and the set $S$ is defined by

$$D^w(p,S) = \sum_{q \in S} w_q\big(|x_p - x_q| + |y_p - y_q|\big), \tag{1}$$

where the discrete point $q$ in $S$ has weight $w_q$. Eq. (1) makes sense since the discrete point $q$ in $S$ with weight $w_q$ can be viewed as $w_q$ points at the same location $(x_q, y_q)$, each point with unit weight. Therefore, the element of medians to be determined, say $m$, for $S$ must satisfy

$$D^w(m,S) = \min_{p \in \mathbb{Z}^2} \sum_{q \in S} w_q\big(|x_p - x_q| + |y_p - y_q|\big). \tag{2}$$

As shown in Fig. 1, suppose we are given a set $S$ consisting of 6 discrete weighted points with weights 3, 3, 1, 2, 2, and 1. The smallest rectangle $R(S)$ containing $S$ is of area 36 $(= 6 \times 6)$.

In general, suppose $S$ is contained in a rectangle $R(S)$ with area $m \times n$. Following the definition of horizontal and vertical projections used in [1,4], let the $i$th vertical projection of $R(S)$, $0 \leqslant i \leqslant n-1$, be denoted by $v_i^w$ and $v_i^w$ is the sum of the weights in the $i$th column. Let the $j$th horizontal projection of $R(S)$, $0 \leqslant i \leqslant m-1$, be denoted by $h_j^w$ and $h_j^w$ is the sum of the weights in the $j$th row. The vertical (horizontal) projection of $R(S)$ is denoted by a vector

$$V^w = \big(v_0^w, v_1^w, \ldots, v_{n-1}^w\big)$$

$(H^w = (h_0^w, h_1^w, \ldots, h_{m-1}^w))$. Return to Fig. 1. We thus have $V = (2,3,0,2,0,5)$ and $H = (1,2,6,0,0,3)$.

We further define the prefix values of $V^w = (v_0^w, v_1^w, \ldots, v_{n-1}^w)$ (respectively $H^w = (h_0^w, h_1^w, \ldots, h_{m-1}^w)$) by $V_0^p = 0$ and $V_i^p = \sum_{k=0}^{i} v_k^w$ for $i = 0, \ldots, n-1$ (respectively $H_0^p = 0$ and $H_j^p = \sum_{k=0}^{j} h_k^w$ for $i = 0, \ldots, m-1$). The total weight of $S$ is defined by

$$W = \sum_{i=0}^{n-1} v_i^w = \sum_{j=0}^{m-1} h_j^w.$$

In Fig. 1, the total weight is $W = 12$ and we have $(V_0^p, V_1^p, V_2^p, V_3^p, V_4^p, V_5^p) = (2,5,5,7,7,12)$ and $(H_0^p, H_1^p, H_2^p, H_3^p, H_4^p, H_5^p) = (1,3,9,9,9,12)$.

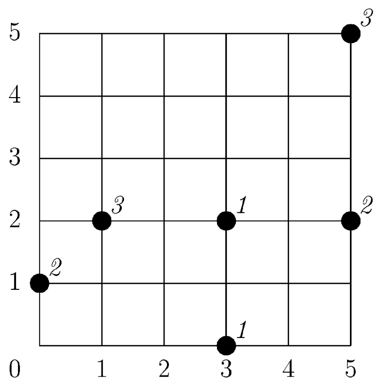

Fig. 1. An example for $S$ and $R(S)$.

## 3. The main result

Following the similar notations used in [4], in our weighted case, the $i$th column of $R(S)$ is said to be a weighted median column if $V_{i-1}^p \leqslant W/2 \leqslant V_i^p$ for $0 \leqslant i \leqslant n-1$; the $j$th row of $R(S)$ is said to be a weighted median row if $H_{j-1}^p \leqslant W/2 \leqslant H_j^p$.

Let $M^w(S)$ be the set of medians for $S$ and following the proving technique used in [4], the next lemma extends the result of Del Lungo et al. [4] to the weighted case.

**Lemma 1.** *A position $M \in M^w(S)$ if and only if $M$ is the intersection of a weighted median column and a weighted median row of $R(S)$.*

**Proof.** First, we prove the only if part of the theorem. Suppose $M = (x_i, y_j)$ belongs to $M^w(S)$. Considering the adjacent point of $M$, say $p = (x_p, y_p)$, we first discuss the horizontal case: $x_p = x_i - 1$ and $y_p = y_j$. From Eq. (2), we have

$$\begin{aligned} D^w(p,S) &- D^w(M,S) \\ &= \sum_{r \in S} w_r\big(|x_p - x_r| + |y_p - y_r|\big) \\ &\quad - \sum_{r \in S} w_r\big(|x_i - x_r| + |y_j - y_r|\big). \end{aligned} \tag{3}$$

Since $y_j = y_p$, it yields $|y_p - y_r| = |y_j - y_r|$. From $x_p = x_i - 1$, we have

$$|x_p - x_r| = \begin{cases} x_i - x_r - 1 & \text{if } x_r \leqslant x_p, \\ x_r - x_i + 1 & \text{otherwise.} \end{cases}$$

From Eq. (3), representing $x_i$ and $y_i$ in terms of $x_p$ and $y_p$, respectively, it yields

$$\begin{aligned} &D^w(p, S) - D^w(M, S) \\ &\quad = \sum\{w_r \mid x_r > x_p,\ r \in S\} \\ &\qquad - \sum\{w_r \mid x_r \leqslant x_p,\ r \in S\}. \end{aligned} \tag{4}$$

From the definition of medians of $S$, since $M$ is a median, we have $D^w(p, S) - D^w(M, S) \geqslant 0$. From Eq. (4), it follows that

$$\begin{aligned} &D^w(p, S) - D^w(M, S) \\ &\quad = \sum\{w_r \mid x_r > x_p,\ r \in S\} \\ &\qquad - \sum\{w_r \mid x_r \leqslant x_p,\ r \in S\} \\ &\quad = W - V^w_{i-1} - V^w_{i-1} \geqslant 0. \end{aligned} \tag{5}$$

We thus have $W/2 \geqslant V^w_{i-1}$.

Under the same horizontal case, we consider another adjacent point of $M$, i.e., $x_p = x_i + 1$ and $y_q = y_j$. By the similar argument as in Eq. (5), it follows that $V^w_i \geqslant W/2$. Combining the result $W/2 \geqslant V^w_{i-1}$, we have $V^w_{i-1} \leqslant W/2 \leqslant V^w_i$.

Considering the vertical adjacent point of $M$, it is not hard to derive that $H^w_{j-1} \leqslant W/2 \leqslant H^w_j$. We have completed the only if part of the theorem.

Finally, we prove the if part of the theorem. Suppose $M = (x_i, y_j)$ is the intersection of a median column and a median row of $S$. We want to prove that $M \in M^w(S)$. From the premise, we have

$$\begin{aligned} &0 \leqslant V^w_1 \leqslant V^w_2 \leqslant \cdots \leqslant V^w_{i-1} \leqslant W/2 \\ &\quad \leqslant V^w_i \leqslant \cdots \leqslant V^w_{m-1} \leqslant V^w_m \end{aligned}$$

and

$$\begin{aligned} &0 \leqslant H^w_1 \leqslant H^w_2 \leqslant \cdots \leqslant H^w_{j-1} \leqslant W/2 \\ &\quad \leqslant H^w_j \leqslant \cdots \leqslant H^w_{n-1} \leqslant H^w_n. \end{aligned}$$

Following the lattice path proving technique used in [4], the if part of the theorem can be proved. For saving space, we omit the detailed derivation. We complete the proof. □

From Lemma 1, it is known that it takes $O(|R(S)|)$ time to find the median of $S$. We first calculate the horizontal and vertical projections using $O(|R(S)|)$ time. Next, from the intersection of a weighted median column and a weighted median row, the time complexity is bounded by $O(|R|)$ to find the median.

In Fig. 1, it is known that $V^w = (2, 3, 0, 2, 0, 5)$, $H^w = (1, 2, 6, 0, 0, 3)$, and $W/2 = 6$. The prefix values of $V^w$ and $H^w$ are denoted by $(V^p_0, V^p_1, V^p_2, V^p_3, V^p_4, V^p_5) = (2, 5, 5, 7, 7, 12)$ and $(H^p_0, H^p_1, H^p_2, H^p_3, H^p_4, H^p_5) = (1, 3, 9, 9, 9, 12)$, respectively. Since $5 = V^p_2 \leqslant 6 \leqslant V^p_3 = 7$ and $3 = H^p_1 \leqslant 6 \leqslant H^p_2 = 9$, the found median is the point with coordinate $(3, 2)$.

Suppose the weighted discrete points set $S$ is stored in a row-major order. For example, the set $S$ in Fig. 1 is stored by $\langle (5, 5), (1, 2), (3, 2), (5, 2), (0, 1), (3, 0) \rangle$. Using the sparse matrix representation technique, the input set $S$ can be represented by a linked list data structure. Based on the row-major scanning order, it takes $O(|S|)$ time to construct the corresponding linked list representation for $S$.

In our example, the vertical projection of $S$ can be represented by

$$V^{sparse} = \left(v^w_0, v^w_1, v^w_3, v^w_5\right) = (2, 3, 2, 5);$$

the horizontal projection of $S$ can be represented by

$$H^{sparse} = \left(h^w_0, h^w_1, h^w_2, h^w_5\right) = (1, 2, 6, 3).$$

The prefix values of $V^{sparse}$ are denoted by $(V^p_0, V^p_1, V^p_3, V^p_5) = (2, 5, 7, 12)$ and the prefix values of $H^{sparse}$ are denoted by $(H^p_0, H^p_1, H^p_2, H^p_5) = (1, 3, 9, 12)$. It is clear that it takes $O(|S|)$ time to construct the vertical/horizontal projections and the corresponding prefix values.

From Lemma 1 and using the sparse matrix representation technique described in the previous two paragraphs, we have the main result.

**Theorem 1.** *Given a weighted discrete points set $S$, the problem of finding the median of $S$ can be solved in $O(|S|)$ time.*

## References

[1] E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Reconstructing convex polyominoes from their horizontal and vertical projections, Theoret. Comput. Sci. 155 (1996) 321–347.

[2] V. Chepoi, F. Dragan, Computing a median point of a simple rectilinear polygon, Inform. Process. Lett. 49 (1994) 281–285.
[3] Y. Métivier, N. Saheb, Medians and centres of polyominoes, Inform. Process. Lett. 57 (1996) 175–181.
[4] A. Del Lungo, M. Nivat, R. Pinzani, L. Sorri, The medians of discrete sets, Inform. Process. Lett. 65 (1998) 293–299.