An approximation algorithm for the continuous k-medians problem in a convex polygon

John Gunnar Carlsson, Fan Jia, and Ying Li

February 3, 2012

Abstract

We give a fast and simple factor 2.74 approximation algorithm for the problem of choosing the k medians of the continuum of demand points defined by a convex polygon C. Our algorithm first surrounds the input region with a bounding box, then subdivides the bounding box into subregions with equal area. Simulation results on the convex hulls of the 50 states in the USA show that the practical performance of our algorithm is within 10% of the optimal solution in the vast majority of cases.

1 Introduction

The k-medians problem, also called the *multi-source Weber problem* [3], is a well-studied geometry problem where the objective is to select a set of k "landmark" points so as to minimize the total distance between the landmark points and some other set of "client" points. The most natural setting for this problem is to let the clients be a discrete set of points in a the plane, which was proved to be NP-hard in [8]. The papers [2] and [7] both describe PTASes for this case. In a general metric space, [4] describes a factor $6\frac{2}{3}$ approximation algorithm.

Another setting is the case where client points form a continuum. This is a natural problem in facility location; in many such scenarios the number of clients is large (say over 1000), which makes the corresponding discrete k-medians problem intractable. Moreover, it may be more sensible, from the standpoint of modelling, to think of clients as being continuously distributed. The first exact algorithmic study for this problem was performed in [5], which describes polynomial-time algorithms for various versions of the 1-median (Fermat-Weber) problem under the L_1 norm. The authors also consider the multiple-center version of the L_1 k-median problem, which they prove is NP-hard for large k. It is also possible to obtain a PTAS to this problem by discretizing the region in question into grid cells and then applying one of the previously mentioned PTASes; however, the running time of such a discretization depends on the "fatness" of the input shape, because a long and skinny input region will require more grid cells to obtain a sufficiently refined grid approximation. In addition, as of this writing, we are unaware of any implementations of the PTASes for the discrete case of our problem, likely due to their fairly sophisticated nature (and possibly poor practical running time).

In this paper, we give a very simple constant-factor approximation algorithm for the continuous k-medians problem in a convex polygon C with n vertices under the L_2 norm. A worst-case theoretical analysis shows that our algorithm always produces solutions within a factor of 2.74 of optimality. In addition, simulation results applied to the convex hulls of the 50 states of the USA show that our algorithm generally performs within 10% of optimality in practice.

Preliminaries The notational conventions of this paper are as follows: we define

$$\begin{aligned} \mathrm{FW}(C) &= \min_p \iint_C \|x-p\|\, dA \\ \mathrm{FW}(C,k) &= \min_{P:|P|=k} \iint_C \min_i \|x-p_i\|\, dA \end{aligned}$$

to be the Fermat-Weber objective functions that we seek to minimize, where $\|\cdot\|$ denotes the Euclidean norm. Let $\Box C$ denote the *minimum-area bounding box* of C (which can be computed in linear running time [10]), and let $\mathrm{width}(C)$ and $\mathrm{height}(C)$ denote the dimensions of $\Box C$. Let $\mathrm{AR}(C)$ denote the *aspect ratio* of C, $\max\left\{\frac{\mathrm{width}(C)}{\mathrm{height}(C)}, \frac{\mathrm{height}(C)}{\mathrm{width}(C)}\right\}$.

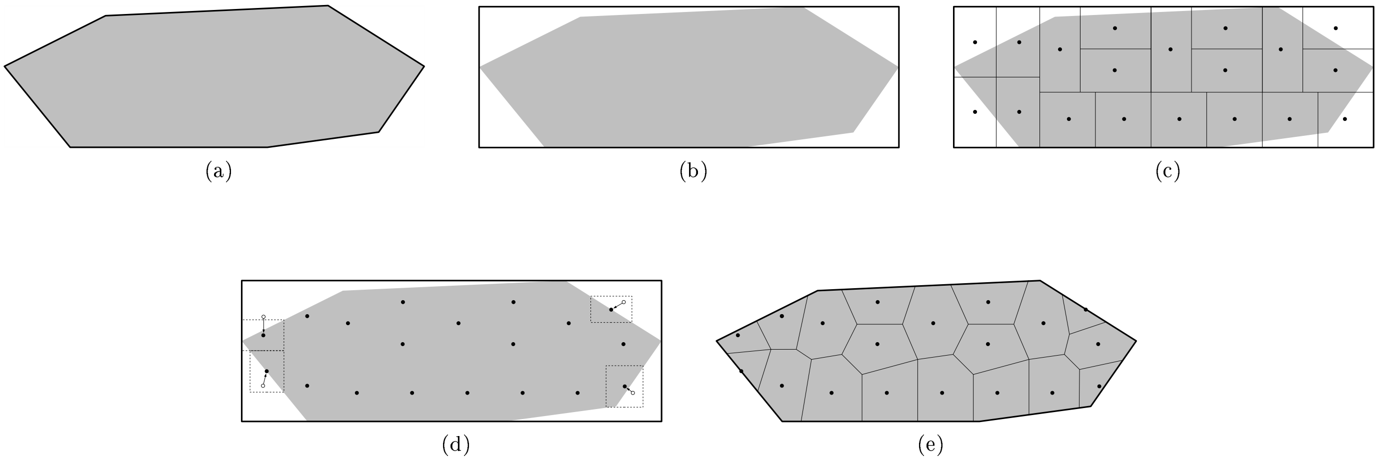


Figure 1: The input and output of Algorithm 2. We begin in (1a) with a convex polygon C, whose minimum-area bounding rectangle $\square C$ is computed in (1b). The bounding box is then partitioned into equal-area pieces in (1c) using Algorithm 1. Some of the centers of these pieces are then relocated in (1d), and (1e) shows the output and Voronoi partition.

2 The algorithm

The input to our algorithm is a convex polygon C with n vertices and an integer k. We assume without loss of generality that C is aligned so that the long side of $\square C$ is aligned with the coordinate x-axis. Note that by construction, it must be the case that $\text{Area}(\square C)/2 \leq \text{Area}(C) \leq \text{Area}(\square C)$. Then, we let $k_1 = \lfloor k/2 \rfloor$ and $k_2 = \lceil k/2 \rceil$ and divide $\square C$ into two pieces of areas $\frac{k_1}{k} \cdot \text{Area}\left(\square C\right) = \frac{k_1}{k} \cdot wh$ and $\frac{k_2}{k} \cdot \text{Area}\left(\square C\right) = \frac{k_2}{k} \cdot wh$ respectively, using a vertical line. This is performed recursively (with the option to split using a horizontal line, if the height of an intermediate rectangle exceeds its width) until all regions have area $\text{Area}\left(\square C\right)/k$. This is given in Algorithms 1 and 2 and Figure 1.

The following lemma is a simplified restatement of a result from [1]:

Lemma 1. *Suppose that $\tilde{R} \subseteq \square C$ is an intermediate rectangle obtained throughout Algorithm 1, which is further subdivided into $\tilde{R}^{'}$ and $\tilde{R}^{''}$. Then:*

1. *If $\text{AR}(\tilde{R}) > 3$, then*
$$\text{AR}(\tilde{R}^{'}), \text{AR}(\tilde{R}^{''}) \leq \text{AR}(\tilde{R})\,.$$
2. *If $\text{AR}(\tilde{R}) \leq 3$, then*
$$\text{AR}(\tilde{R}^{'}), \text{AR}(\tilde{R}^{''}) \leq 3\,.$$

Proof. Claim 1 is trivial. To prove claim 2 we assume that $\text{AR}(\tilde{R}) \leq 3$. Assume without loss of generality that $\text{width}(\tilde{R}) \geq \text{height}(\tilde{R})$, so that $\text{height}(\tilde{R}^{'}) = \text{height}(\tilde{R})$. Since $\tilde{R}$ is always divided into proportions as close as $1/2$ as possible, we have

$$\text{width}(\tilde{R})/3 \leq \text{width}(\tilde{R}^{'}) \leq 2\,\text{width}(\tilde{R})/3$$

and, dividing by $\text{height}(\tilde{R})$, we find that

$$\text{width}(\tilde{R})/(3\,\text{height}(\tilde{R})) \leq \text{width}(\tilde{R}^{'})/\,\text{height}(\tilde{R}^{'}) = \text{width}(\tilde{R}^{'})/\,\text{height}(\tilde{R}) \leq 2\,\text{width}(\tilde{R})/(3\,\text{height}(\tilde{R})) \leq 2$$

so that $\text{width}(\tilde{R}^{'})/\,\text{height}(\tilde{R}^{'}) \leq 2$. Taking the reciprocal of this expression and observing that $3 \geq 3\,\text{height}(\tilde{R})/\,\text{width}(\tilde{R})$ since $\text{width}(\tilde{R}) \geq \text{height}(\tilde{R})$, we have

$$3 \geq 3\,\text{height}(\tilde{R})/\,\text{width}(\tilde{R}) \geq \text{height}(\tilde{R}^{'})/\,\text{width}(\tilde{R}^{'}) = \text{height}(\tilde{R})/\,\text{width}(\tilde{R}^{'}) \geq 3\,\text{height}(\tilde{R})/(2\,\text{width}(\tilde{R}))$$

so that $3 \geq \text{height}(\tilde{R}^{'})/\,\text{width}(\tilde{R}^{'})$. This same argument clearly applies to $\tilde{R}^{''}$ as well, which completes claim 2. $\square$

Input: An axis-aligned rectangle R and an integer k.
Output: A partition of R into k rectangles, each having area $\text{Area}(R)/k$.
if $k = 1$ **then**
  return R;
else
  Set $k_1 = \lfloor k/2 \rfloor$ and $k_2 = \lceil k/2 \rceil$;
  Let w denote the width of R and h the height;
  if $w \geq h$ **then**
    With a vertical line, divide R into two pieces R_1 and R_2 with area $\frac{k_1}{k} \cdot \text{Area}(R)$ on the right and $\frac{k_2}{k} \cdot \text{Area}(R)$ on the left;
  else
    With a horizontal line, divide R into two pieces R_1 and R_2 with area $\frac{k_1}{k} \cdot \text{Area}(R)$ on the top and $\frac{k_2}{k} \cdot \text{Area}(R)$ on the bottom;
  end
  return $\mathit{RectanglePartition}(R_1, k_1) \cup \mathit{RectanglePartition}(R_2, k_2)$;
end

Algorithm 1: Algorithm RectanglePartition (R, k) takes as input an axis-aligned rectangle R and a positive integer k.

Input: A convex polygon C and an integer k.
Output: The locations of k points p_i in C that approximately minimize $\text{FW}(C, k)$ within a factor of 2.74.
Let $\square C$ denote a minimal-area bounding box of C;
Rotate C so that $\square C$ is aligned with the coordinate axes;
Let $R_1, \ldots, R_k = \text{RectanglePartition}(\square C, k)$;
for $i \in \{1, \ldots, k\}$ **do**
  Let c_i denote the center of R_i;
  if $c_i \in C$ **then**
    Set $p_i = c_i$;
  else
    if $R_i \cap C$ *is nonempty* **then**
      Let $R_i^{'}$ be the minimum axis-aligned bounding box of $R_i \cap C$ and let $c_i^{'}$ denote its center;
      Set $p_i = c_i^{'}$;
    else
      Place p_i anywhere in C;
    end
  end
end
return $p_1, \ldots, p_k$;

Algorithm 2: Algorithm ApproxFW (R, k) takes as input a convex polygon C and an integer k.

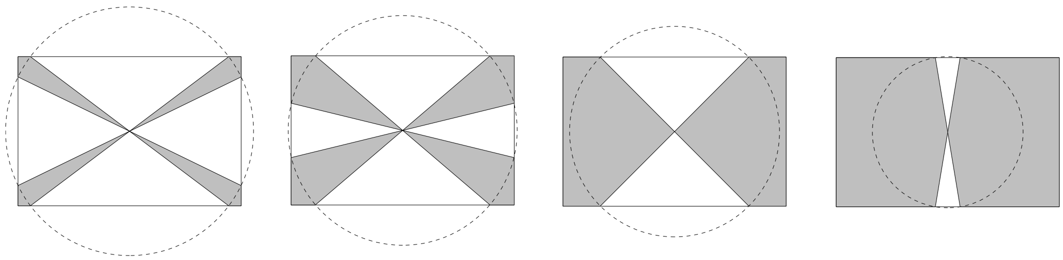


Figure 2: The optimal regions C^* in a given box B, for increasing values of A.

3 Upper and lower bounds

In this section we establish some upper and lower bounds for the Fermat-Weber value of any convex region C. We begin with some simple lemmas:

Lemma 2. *For a disk D with radius r,*

$$\mathrm{FW}(D) = \frac{2\pi r^3}{3}.$$

Proof. Trivial. □

Remark 3. It is well-known that, for a fixed area, the disk is the region with minimal Fermat-Weber value $\mathrm{FW}(C)$. This gives us an easy lower bound:

$$\mathrm{FW}(C) \geq \frac{2}{3\sqrt{\pi}} A^{3/2}$$

where A is the area of C.

Definition 4. A region C is said to be *star convex at the point* p if the line segment from p to any point $x \in C$ is itself contained in C. Similarly, the *star convex hull* of a region S at the point p is the smallest star-convex region at the point p that contains S (i.e. the union of all segments between points $x \in S$ and p).

Lemma 5. *Let B be a box of dimensions $w \times h$ centered at the origin. The region C^* that solves the infinite-dimensional optimization problem*

$$\begin{aligned} \underset{C}{\text{maximize}}\, \mathrm{FW}(C) & \quad & s.t. \\ C & \subseteq & B \\ \mathrm{Area}(C) & = & A \\ C & \ni & (0,0) \\ C & \text{is} & \text{star convex at } (0,0) \end{aligned} \tag{1}$$

is the star convex hull of $B \backslash D$, where D is an appropriately chosen disk centered at the origin, as indicated in Figure 2. Furthermore for fixed w and h, the function $\Phi(A) = \mathrm{FW}(C^*)$ (i.e. the maximal value of (1)) is monotonically increasing and concave.

Proof sketch. This follows from a standard argument where we consider the integer (or linear) program obtained by discretizing problem (1) using polar coordinates. See Figure 3. Concavity of $\Phi(A)$ follows by observing that we build our optimal solution by adding sectors containing points that are strictly closer than the points in the sector that preceded them. □

Definition 6. Let $\mathbf{x} = (x, y)$ be a point in the plane. We define the $\boxtimes$ and ⧇ norms by

$$\begin{aligned} \|\mathbf{x}\|_{\boxtimes} &= \max\{|x|, |y|\} + \left(\sqrt{2}-1\right)\min\{|x|, |y|\} \\ \|\mathbf{x}\|_{⧇} &= \max\left\{|x|, |y|, \frac{1}{\sqrt{2}}(|x|+|y|)\right\}. \end{aligned}$$

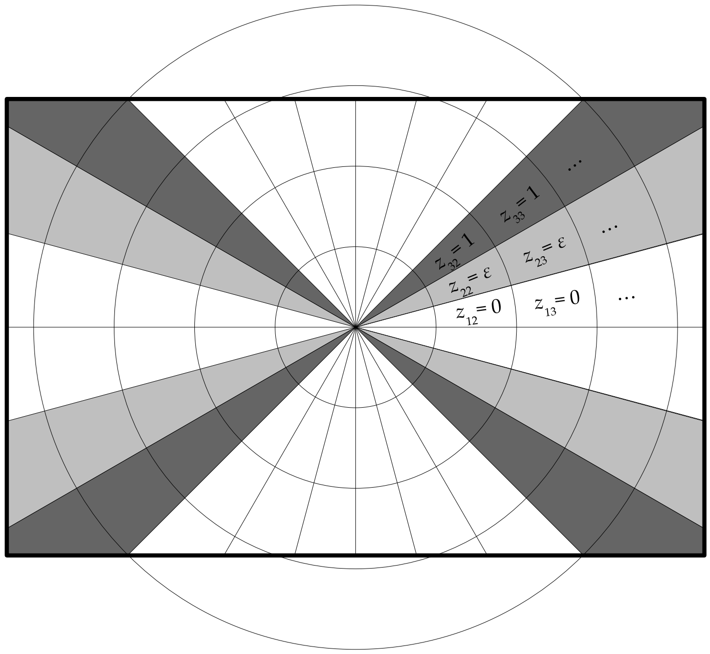


Figure 3: In the discretization above, our variables are set up in such a way that the star convexity constraint is equivalent to setting $z_{i(j+1)} \leq z_{ij}$ for all j. By the nature of the constraints it is clear that we may assume that $z^*_{i(j+1)} = z^*_{ij}$ at optimality since the distance from cell ij to the origin increases with j. The diagram above suggests a linear programming formulation, where the lighter regions indicate fractional solutions.

Remark 7. The following identities are easy to verify:

$$\begin{aligned} \|\mathbf{x}\|_{\boxdiamond} &\leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_{\boxtimes} \\ \psi \|\mathbf{x}\|_{\boxtimes} &\leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_{\boxtimes} \\ \|\mathbf{x}\|_{\boxdiamond} &\leq \|\mathbf{x}\|_2 \leq \frac{1}{\psi} \|\mathbf{x}\|_{\boxdiamond} \end{aligned}$$

where $\psi = \cos(\pi/8) \approx 0.9239$. Both norms have a natural interpretation: $\|\mathbf{x}\|_{\boxtimes}$ is the distance from $(0,0)$ to $\mathbf{x}$ if we are only permitted to move horizontally, vertically, or diagonally (the cardinal and ordinal directions) and $\|\mathbf{x}\|_{\boxdiamond}$ is the maximum distance from $(0,0)$ to $\mathbf{x}$ in the horizontal, vertical, or diagonal direction.

Lemma 8. *Let C be a convex region, contained in a box B of dimensions $w \times h$, that contains the center $(0,0)$ of B. If $A = \text{Area}(C)$, then we have*

$$\text{FW}(C) \leq \begin{cases} \frac{\sqrt{2}-1}{3} Ah - \frac{\sqrt{2}-1}{12} wh^2 - \frac{\sqrt{w^2-h^2}}{3}(wh - A) + \frac{1}{3}w^2h - \frac{1}{12}h^3 & \text{if } A \leq wh - \frac{h}{2}\sqrt{w^2-h^2} \\ \frac{2}{3}Aw + \frac{\sqrt{2}-1}{3}Ah - \frac{1}{3} \cdot \frac{A^2}{h} - \frac{1}{12}w^2h - \frac{\sqrt{2}-1}{12}wh^2 & \text{if } wh - \frac{h}{2}\sqrt{w^2-h^2} < A \leq wh - h^2/2 \\ \frac{2\sqrt{2}-2}{3}Aw + \frac{1}{3}Ah + \frac{7-4\sqrt{2}}{12}w^2h + \frac{2-\sqrt{2}}{12}h^3 - \frac{\sqrt{2}-1}{3} \cdot \frac{A^2}{h} - \frac{7-3\sqrt{2}}{12}wh^2 & \text{otherwise} \end{cases} .$$

Proof. We consider the relaxation of the infinite-dimensional optimization problem

$$\begin{aligned} \underset{C}{\text{maximize}}\, \text{FW}(C) \quad & s.t. \\ C &\subseteq B \\ \text{Area}(C) &= A \\ C &\ni (0,0) \\ C &\text{ is convex}, \end{aligned}$$

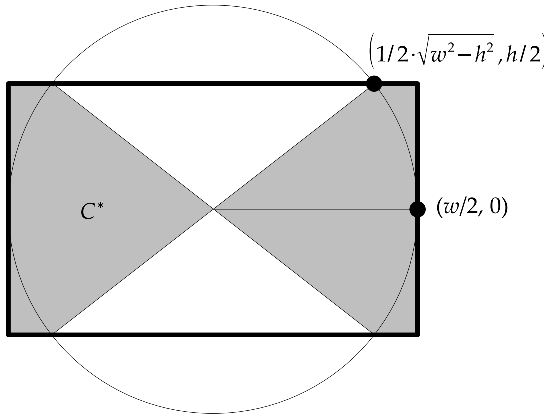


Figure 4: The area of the shaded region is $wh - \frac{h}{2}\sqrt{w^2 - h^2}$.

obtained by replacing the convexity constraint with star convexity about the origin. The problem is now equivalent to problem (1).

Following lemma 5 we see that the optimal star-convex region C^* takes the form shown in Figure 2. If $A \geq wh - \frac{h}{2}\sqrt{w^2 - h^2}$, then the optimal solution consists of two components (rather than 4) as shown in Figure 4. We evaluate the Fermat-Weber objective value under the ⊠ norm and we find that

$$\mathrm{FW}_{\boxtimes}(C^*) = \begin{cases} 4\int_0^{h/2}\int_{y/m}^{w/2} x + \left(\sqrt{2}-1\right) y \, dx \, dy & \text{if } wh - \frac{h}{2}\sqrt{w^2-h^2} \leq A \leq wh - h^2 \\ 4\int_0^{h/2}\int_{y/m}^{y} \left(\sqrt{2}-1\right) x + y \, dx + \int_{y}^{w/2} x + \left(\sqrt{2}-1\right) y \, dx \, dy & \text{if } A > wh - h^2 \end{cases}$$

where $m = \frac{h^2}{2(wh-A)}$; the formulas are thus found by analytic integration. We find the upper bound for the case $A \leq wh - \frac{h}{2}\sqrt{w^2 - h^2}$ simply by using the fact that $\mathrm{FW}_{\boxtimes}(C^*)$ is concave in A, then taking a tangent line at $A = wh - \frac{h}{2}\sqrt{w^2 - h^2}$ and extrapolating. □

For the remainder of this paper we define $\rho = 4.34818$ and $\gamma = 0.11719$.

Lemma 9. *Let C be a convex region with area A, contained in a box B of dimensions $w \times h$, where $w/h \geq \rho$. If $A \geq wh/2$, then*

$$\mathrm{FW}(C) \geq \begin{cases} \frac{\left(16+12\sqrt{2}\right)\gamma(\gamma-1)+\sqrt{2}}{24\left(\sqrt{2}+1\right)^2} h^3 + \frac{8+6\sqrt{2}-\left(28+20\sqrt{2}\right)\gamma}{24\left(\sqrt{2}+1\right)^2} Ah + \frac{16+11\sqrt{2}}{24\left(\sqrt{2}+1\right)^2} A^2/h & \textit{if } A \leq \frac{\left(2\sqrt{2}-2\right)\gamma+1}{\sqrt{2}-1} h^2 \\ \left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) h^3 - (\gamma/2)\, Ah + (1/4)\, A^2/h & \textit{otherwise} \end{cases}.$$

Proof. Refer to Figure 5 for this proof. The shape C^* that minimizes $\mathrm{FW}(C)$ in B is the intersection of a disk with a slab of height h. Let R_1 denote the largest rectangle contained in C^* and let R_2 denote the smallest rectangle containing C^*. Clearly, for fixed h, the two rectangles become the same as $\mathrm{Area}(C^*)$ increases. In particular, if $\mathrm{Area}(C^*) \geq wh/2 \geq \rho h^2/2$, then we can verify numerically that the "gap" g between the two (as indicated in Figure 5b) is at most γh. Therefore, the hexagon with vertices $(\pm A/2h, 0)$ and $(\pm(A/2h - \gamma h), \pm h/2)$ is contained entirely in C^* whenever $\mathrm{Area}(C^*) \geq \rho h^2/2$. To obtain the desired result, we bound the Fermat-Weber value on this hexagon under the ⊠ norm:

$$\mathrm{FW}_{\boxtimes}(C^*) \geq \begin{cases} 4\int_0^{h/2}\int_0^{y/m_1} y \, dx + \frac{1}{\sqrt{2}}\int_{y/m_1}^{y/m_2} x + y \, dx + \int_{y/m_2}^{-2\gamma y + A/2h} x \, dx \, dy & \text{if } A \leq \frac{\left(2\sqrt{2}-2\right)\gamma+1}{\sqrt{2}-1} h^2 \\ 4\int_0^{h/2}\int_0^{y/m_1} y \, dx + \frac{1}{\sqrt{2}}\int_{y/m_1}^{y/m_3} x + y \, dx + \int_{y/m_3}^{-2\gamma y + A/2h} x \, dx \, dy & \text{otherwise} \end{cases}$$

where $m_1 = \sqrt{2} + 1$, $m_2 = \frac{h^2}{A - 2\gamma h^2}$, and $m_3 = \sqrt{2} - 1$, which gives the desired bounds. □

In summary, we have the following upper and lower bounds for $\mathrm{FW}(C)$ when C is a convex region of area A contained in a box of dimensions $w \times h$:

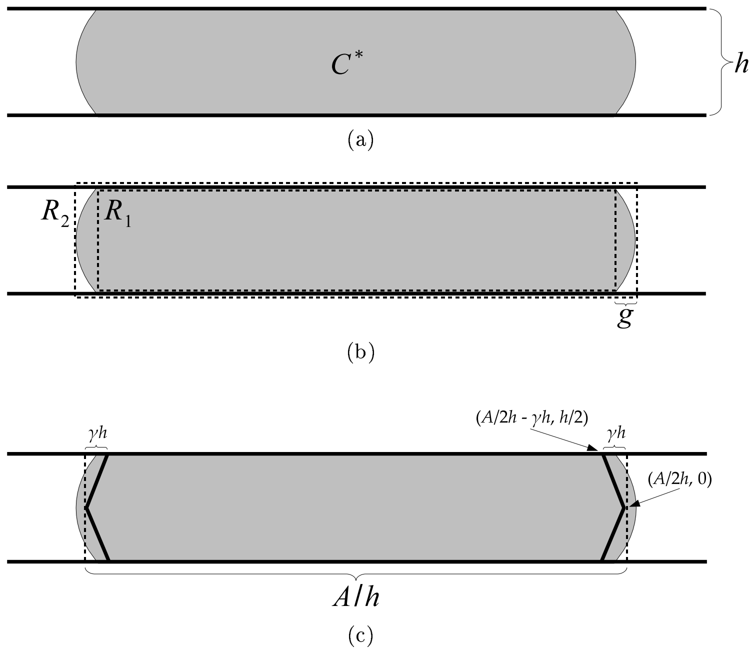


Figure 5: The region C^* of minimal Fermat-Weber objective value, the rectangles R_1 and R_2, and the hexagon that gives our lower bound.

$$\mathrm{FW}(C) \leq \begin{cases} \frac{\sqrt{2}-1}{3}Ah - \frac{\sqrt{2}-1}{12}wh^2 - \frac{\sqrt{w^2-h^2}}{3}(wh-A) + \frac{1}{3}w^2h - \frac{1}{12}h^3 & \text{if } A \leq wh - \frac{h}{2}\sqrt{w^2-h^2} \\ \frac{2}{3}Aw + \frac{\sqrt{2}-1}{3}Ah - \frac{1}{3}\cdot\frac{A^2}{h} - \frac{1}{12}w^2h - \frac{\sqrt{2}-1}{12}wh^2 & \text{if } wh - \frac{h}{2}\sqrt{w^2-h^2} < A \leq wh - h^2/2 \\ \frac{2\sqrt{2}-2}{3}Aw + \frac{1}{3}Ah + \frac{7-4\sqrt{2}}{12}w^2h + \frac{2-\sqrt{2}}{12}h^3 - \frac{\sqrt{2}-1}{3}\cdot\frac{A^2}{h} - \frac{7-3\sqrt{2}}{12}wh^2 & \text{otherwise} \end{cases} \quad (2)$$

$$\mathrm{FW}(C) \geq \frac{2}{3\sqrt{\pi}}A^{3/2} \quad (3)$$

$$\mathrm{FW}(C) \geq \begin{cases} \frac{\left(16+12\sqrt{2}\right)\gamma(\gamma-1)+\sqrt{2}}{24\left(\sqrt{2}+1\right)^2}h^3 + \frac{8+6\sqrt{2}-\left(28+20\sqrt{2}\right)\gamma}{24\left(\sqrt{2}+1\right)^2}Ah + \frac{16+11\sqrt{2}}{24\left(\sqrt{2}+1\right)^2}A^2/h & \text{if } A \leq \frac{\left(2\sqrt{2}-2\right)\gamma+1}{\sqrt{2}-1}h^2 \\ \left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right)h^3 - (\gamma/2)\,Ah + (1/4)\,A^2/h & \text{otherwise} \end{cases} \quad \text{if } w/h \geq \rho\,. \quad (4)$$

Notice that both of our lower bounds are convex in A.

4 Proof of approximation

In this section, we show that Algorithm 2 has a constant-factor approximation. After performing our algorithm, we have a collection of k rectangles with area wh/k; if $k \geq w/\rho h$, then these rectangles have an aspect ratio not exceeding ρ. If $k < w/\rho h$, then all rectangles are identical and have dimensions $(w/k) \times h$. Since our lower bounds are convex in A and the function $\Phi(A)$ is concave, we immediately know that the worst possible ratio between the upper and lower bounds for $\mathrm{FW}(C, k)$ is attained when all rectangles contain area A/k of C. Therefore, to find the approximation bounds for this algorithm, it will suffice to consider a single such rectangle with height 1 and width $z \geq 1$, containing area $\alpha = A/k$ of C. The approximation bounds are determined by the value of A and the relationship between k and $w/\rho h$. Our general approach is to reduce the approximation ratio to a function of a single variable, whose upper bound can be easily be verified using standard methods from calculus. We omit these steps for brevity.

4.1 The case $k \le w/\rho h$

If $k \le w/\rho h$, then our algorithm will divide $\square C$ into k identical rectangles of dimensions $(w/k) \times h$. Thus we set $z = w/k \ge \rho$ and assume without loss of generality that $h = 1$. We will use lower bound (4). Note that we may either have $\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}h^2 < wh - \frac{h}{2}\sqrt{w^2-h^2}$ or $\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}h^2 \ge wh - \frac{h}{2}\sqrt{w^2-h^2}$, depending on the dimensions of $\square C$, and therefore we have to consider these cases separately. In total, there are five cases that we have to consider:

1. $A \le wh - \frac{h}{2}\sqrt{w^2-h^2}$ and $A \le \frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}h^2$.
2. $\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}h^2 < A \le wh - \frac{h}{2}\sqrt{w^2-h^2}$.
3. $wh - \frac{h}{2}\sqrt{w^2-h^2} < A \le \frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}h^2$.
4. $A > wh - \frac{h}{2}\sqrt{w^2-h^2}$, $A > \frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}h^2$, and $A \le wh - h^2/2$.
5. $A > wh - h^2/2$.

We consider each case separately below.

Case 1 By assumption we have $\alpha \le z - \frac{1}{2}\sqrt{z^2-1}$ and $\alpha \le \frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}$ and the approximation ratio is given by

$$\frac{\frac{\sqrt{2}-1}{3}\alpha - \frac{\sqrt{2}-1}{12}z - \frac{\sqrt{z^2-1}}{3}(z-\alpha) + \frac{1}{3}z^2 - \frac{1}{12}}{\frac{(16+12\sqrt{2})\gamma(\gamma-1)+\sqrt{2}}{24(\sqrt{2}+1)^2} + \frac{8+6\sqrt{2}-(28+20\sqrt{2})\gamma}{24(\sqrt{2}+1)^2}\alpha + \frac{16+11\sqrt{2}}{24(\sqrt{2}+1)^2}\alpha^2}.$$

We notice that the denominator is quadratic in α and the numerator is linear in α, and it is therefore not hard to show that the ratio is maximized when α is as small as possible, i.e. that $\alpha = z/2$. Since we have $\alpha \le \frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}$, it must be the case that $z \le \frac{(4\sqrt{2}-4)\gamma+2}{\sqrt{2}-1} < 5.30$. The approximation ratio is therefore

$$\frac{\frac{\sqrt{2}-1}{6}z - \frac{\sqrt{2}-1}{12}z - \frac{\sqrt{z^2-1}}{6}z + \frac{1}{3}z^2 - \frac{1}{12}}{\frac{(16+12\sqrt{2})\gamma(\gamma-1)+\sqrt{2}}{24(\sqrt{2}+1)^2} + \frac{8+6\sqrt{2}-(28+20\sqrt{2})\gamma}{48(\sqrt{2}+1)^2}z + \frac{16+11\sqrt{2}}{96(\sqrt{2}+1)^2}z^2}$$

which is bounded above by 2.74 for $z \in [\rho, 5.30]$.

Case 2 By assumption we have $\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1} < \alpha \le z - \frac{1}{2}\sqrt{z^2-1}$ and the approximation ratio is given by

$$\frac{\frac{\sqrt{2}-1}{3}\alpha - \frac{\sqrt{2}-1}{12}z - \frac{\sqrt{z^2-1}}{3}(z-\alpha) + \frac{1}{3}z^2 - \frac{1}{12}}{\left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) - (\gamma/2)\,\alpha + (1/4)\,\alpha^2}.$$

We again observe that the denominator is quadratic in α and the numerator is linear in α, and it is therefore not hard to show that the ratio is maximized when α is as small as possible, i.e. that $\alpha = \max\left\{\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}, z/2\right\}$. The approximation ratio is given by

$$\frac{\frac{\sqrt{2}-1}{3}\max\left\{\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}, z/2\right\} - \frac{\sqrt{2}-1}{12}z - \frac{\sqrt{z^2-1}}{3}\left(z - \max\left\{\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}, z/2\right\}\right) + \frac{1}{3}z^2 - \frac{1}{12}}{\left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) - (\gamma/2)\max\left\{\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}, z/2\right\} + (1/4)\max\left\{\frac{(2\sqrt{2}-2)\gamma+1}{\sqrt{2}-1}, z/2\right\}^2}$$

which is bounded above by 2.74 for $z \ge \rho$.

Case 3 By assumption we have $z - \frac{1}{2}\sqrt{z^2-1} < \alpha \le \frac{\left(2\sqrt{2}-2\right)\gamma+1}{\sqrt{2}-1}$ and the approximation ratio is given by

$$\frac{\frac{2}{3}\alpha z + \frac{\sqrt{2}-1}{3}\alpha - \frac{1}{3}\alpha^2 - \frac{1}{12}z^2 - \frac{\sqrt{2}-1}{12}z}{\frac{\left(16+12\sqrt{2}\right)\gamma(\gamma-1)+\sqrt{2}}{24\left(\sqrt{2}+1\right)^2} + \frac{8+6\sqrt{2}-\left(28+20\sqrt{2}\right)\gamma}{24\left(\sqrt{2}+1\right)^2}\alpha + \frac{16+11\sqrt{2}}{24\left(\sqrt{2}+1\right)^2}\alpha^2}.$$

We notice that the numerator is increasing in z, and therefore we assume that z is as large as possible so that $\alpha = z - \frac{1}{2}\sqrt{z^2-1}$. Furthermore, since $z - \frac{1}{2}\sqrt{z^2-1} < \frac{\left(2\sqrt{2}-2\right)\gamma+1}{\sqrt{2}-1}$, it must be the case that $z < 5.2002$. The approximation ratio is given by

$$\frac{\frac{2}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)z + \frac{\sqrt{2}-1}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right) - \frac{1}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)^2 - \frac{1}{12}z^2 - \frac{\sqrt{2}-1}{12}z}{\frac{\left(16+12\sqrt{2}\right)\gamma(\gamma-1)+\sqrt{2}}{24\left(\sqrt{2}+1\right)^2} + \frac{8+6\sqrt{2}-\left(28+20\sqrt{2}\right)\gamma}{24\left(\sqrt{2}+1\right)^2}\left(z - \frac{1}{2}\sqrt{z^2-1}\right) + \frac{16+11\sqrt{2}}{24\left(\sqrt{2}+1\right)^2}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)^2}$$

which is bounded above by 2.74 for $z \in [\rho, 5.2002]$.

Case 4 By assumption we have $\alpha > z - \frac{1}{2}\sqrt{z^2-1}$, $\alpha > \frac{\left(2\sqrt{2}-2\right)\gamma+1}{\sqrt{2}-1}$, and $\alpha \le z - 1/2$ and the approximation ratio is given by

$$\frac{\frac{2}{3}\alpha z + \frac{\sqrt{2}-1}{3}\alpha - \frac{1}{3}\alpha^2 - \frac{1}{12}z^2 - \frac{\sqrt{2}-1}{12}z}{\left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) - (\gamma/2)\,\alpha + (1/4)\,\alpha^2}.$$

We notice that the numerator is increasing in z in this domain, and therefore we assume that z is as large as possible so that $\alpha = z - \frac{1}{2}\sqrt{z^2-1}$. The approximation ratio is therefore

$$\frac{\frac{2}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)z + \frac{\sqrt{2}-1}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right) - \frac{1}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)^2 - \frac{1}{12}z^2 - \frac{\sqrt{2}-1}{12}z}{\left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) - (\gamma/2)\left(z - \frac{1}{2}\sqrt{z^2-1}\right) + (1/4)\left(z - \frac{1}{2}\sqrt{z^2-1}\right)^2}$$

which is bounded above by 2.74 for $z \ge \rho$.

Case 5 By assumption we have $\alpha > z - 1/2$ and the approximation ratio is given by

$$\frac{\frac{2\sqrt{2}-2}{3}\alpha z + \frac{1}{3}\alpha + \frac{7-4\sqrt{2}}{12}z^2 + \frac{2-\sqrt{2}}{12} - \frac{\sqrt{2}-1}{3}\alpha^2 - \frac{7-3\sqrt{2}}{12}z}{\left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) - (\gamma/2)\,\alpha + (1/4)\,\alpha^2}.$$

We notice that the numerator is increasing in z, and therefore we assume that z is as large as possible so that $\alpha = z - 1/2$. The approximation ratio is therefore

$$\frac{\frac{2\sqrt{2}-2}{3}(z-1/2)\,z + \frac{1}{3}(z-1/2) + \frac{7-4\sqrt{2}}{12}z^2 + \frac{2-\sqrt{2}}{12} - \frac{\sqrt{2}-1}{3}(z-1/2)^2 - \frac{7-3\sqrt{2}}{12}z}{\left(\gamma^3/3 + \sqrt{2}/6 - 1/12\right) - (\gamma/2)(z-1/2) + (1/4)(z-1/2)^2}$$

which is bounded above by 1.4 for $z \ge \rho$.

4.2 The case $k > w/\rho h$

If $k > w/\rho h$, then our algorithm will divide $\square C$ into k rectangles with aspect ratio at most ρ. Thus we assume that we have k rectangles with height 1 and width $z \in [1, \rho]$. We will use lower bound (3). In total, there are three cases that we have to consider:

6. $A \le wh - \frac{h}{2}\sqrt{w^2-h^2}$.
7. $wh - \frac{h}{2}\sqrt{w^2-h^2} < A \le wh - h^2/2$.
8. $A > wh - h^2/2$.

Case 6 By assumption we have $\alpha \le z - \frac{1}{2}\sqrt{z^2-1}$ and the approximation ratio is given by

$$\frac{\frac{\sqrt{2}-1}{3}\alpha - \frac{\sqrt{2}-1}{12}z - \frac{\sqrt{z^2-1}}{3}(z-\alpha) + \frac{1}{3}z^2 - \frac{1}{12}}{\frac{2}{3\sqrt{\pi}}\alpha^{3/2}}.$$

It is not hard to verify that the above ratio is decreasing in α, and therefore we assume that α is as small as possible, i.e. that $\alpha = z/2$. The approximation ratio is therefore

$$\frac{\frac{\sqrt{2}-1}{6}z - \frac{\sqrt{2}-1}{12}z - \frac{\sqrt{z^2-1}}{6}z + \frac{1}{3}z^2 - \frac{1}{12}}{\frac{2}{3\sqrt{\pi}}\left(z/2\right)^{3/2}}$$

which is bounded above by 2.74 for $z \in [1, \rho]$.

Case 7 By assumption we have $z - \frac{1}{2}\sqrt{z^2-1} < \alpha \le z - 1/2$ and the approximation ratio is given by

$$\frac{\frac{2}{3}\alpha z + \frac{\sqrt{2}-1}{3}\alpha - \frac{1}{3}\alpha^2 - \frac{1}{12}z^2 - \frac{\sqrt{2}-1}{12}z}{\frac{2}{3\sqrt{\pi}}\alpha^{3/2}}.$$

We notice that the numerator is increasing in z in this domain and therefore we assume that $\alpha = z - \frac{1}{2}\sqrt{z^2-1}$. The approximation ratio is therefore

$$\frac{\frac{2}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)z + \frac{\sqrt{2}-1}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right) - \frac{1}{3}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)^2 - \frac{1}{12}z^2 - \frac{\sqrt{2}-1}{12}z}{\frac{2}{3\sqrt{\pi}}\left(z - \frac{1}{2}\sqrt{z^2-1}\right)^{3/2}}$$

which is bounded above by 2.71 for $z \in [1, \rho]$.

Case 8 By assumption we have $\alpha > z - 1/2$ and the approximation ratio is given by

$$\frac{\frac{2\sqrt{2}-2}{3}\alpha z + \frac{1}{3}\alpha + \frac{7-4\sqrt{2}}{12}z^2 + \frac{2-\sqrt{2}}{12} - \frac{\sqrt{2}-1}{3}\alpha^2 - \frac{7-3\sqrt{2}}{12}z}{\frac{2}{3\sqrt{\pi}}\alpha^{3/2}}.$$

We notice that the numerator is increasing in z, and therefore we assume that z is as large as possible so that $\alpha = z - 1/2$. The approximation ratio is therefore

$$\frac{\frac{2\sqrt{2}-2}{3}\left(z-1/2\right)z + \frac{1}{3}\left(z-1/2\right) + \frac{7-4\sqrt{2}}{12}z^2 + \frac{2-\sqrt{2}}{12} - \frac{\sqrt{2}-1}{3}\left(z-1/2\right)^2 - \frac{7-3\sqrt{2}}{12}z}{\frac{2}{3\sqrt{\pi}}\left(z-1/2\right)^{3/2}}$$

which is bounded above by 1.8 for $z \in [1, \rho]$.

4.3 Running time

This algorithm can be performed with running time $\mathcal{O}\left(n + k + k\log n\right)$. This is because Algorithm 1 takes $\mathcal{O}\left(k\right)$ operations to partition the rectangle and Algorithm 2 requires $\mathcal{O}\left(n\right)$ operations to find a minimum bounding box of C. The last step of Algorithm 2 consists of moving the center points to C when necessary, which takes $\mathcal{O}\left(k\log n\right)$ operations using a point-in-polygon algorithm [9].

5 Simulation results

In order to determine the practical performance of our algorithm (as opposed to the theoretical worst-case bounds), we applied it to a dataset generated by forming the convex hulls of the 50 states of the USA. In order to improve the practical performance of Algorithm 2, we introduce one small modification, which we describe below; we have avoided putting this modification in the description of Algorithm 2 because it is not clear how to incorporate it into the theoretical upper-bounding procedure.

Modification to Algorithm 2 In executing Algorithm 1 as a subroutine in Algorithm 2, it may be the case that a final rectangle R_i lies entirely outside C. If this is the case, we add rectangle R_i to a list `OUT`. It may also be the case that an *intermediate* (not necessarily final) rectangle $\tilde{R}$ lies entirely *inside* C; in this case, we add rectangle $\tilde{R}$ to a list `IN`. We do not add any intermediate rectangles to `IN` if they are themselves contained in a larger intermediate rectangle that is already contained in `IN`. Finally, it may be the case that a final rectangle R_i is *partially* contained by C; in this case we also add R_i to list `IN`. See Figure 6a.

After executing Algorithm 1 as a subroutine in Algorithm 2, we now have the lists `OUT` and `IN`. The centers of the rectangles of list `OUT` are clearly not helping us, and we now want to find a reasonable way to move them into C. Let the jth rectangle of list `IN` (which may be an intermediate rectangle) be written as $\tilde{R}_j$, let $\tilde{A}_j$ denote the area of $\tilde{R}_j$, and let N_j denote the number of points that would be assigned to $\tilde{R}_j$ (which is 1 if $\tilde{R}_j$ is a final rectangle). We then let

$$\tilde{N}_j := k \cdot \left\lfloor \frac{\tilde{A}_j}{\sum_{q=1}^{|\mathtt{IN}|} \tilde{A}_q} \right\rfloor$$

and we then place $\tilde{N}_j$ points in each rectangle $\tilde{R}_j$. This does not affect our approximation result because $\tilde{N}_j \geq N_j$ and $\sum_j \tilde{N}_j \leq k$. The remaining points are distributed arbitrarily among the rectangles in `IN`. See Figure 6b.

Results and discussion Figure 7 shows the approximation ratios for $k \leq 1000$ when our algorithm is applied to the convex hulls of the 50 states of the USA. As in the proof of the worst-case approximation ratio, we use lower bounds (3) and (4). However, instead of using the theoretical upper bound (2), we simply evaluate the actual objective function value

$$\iint_C \min_i \|x - p_i\| \, dA$$

by first taking a Voronoi diagram of the points p_i and then integrating the distance function $\|x - p_i\|$ over each Voronoi cell using the collapsed-square Gaussian cubature method [6] with tolerance 10^{-5}. As an example, Figure 8 shows the output of our algorithm for $k = 100$ applied to the convex hull of the state of Minnesota. The running times of our trials are basically trivial as explained in Section 4.3 and we have therefore omitted them.

We first observe that in nearly all of the cases (the exceptions being very low values of k), our algorithm gives a solution that is within 10% of the theoretical lower bound; this suggests both that our algorithm generally performs well and that lower bounds (3) and (4) are fairly tight. This also suggests that the upper bound (2) is the weak point of our worst-case performance bounds. It is also interesting that the approximation ratios in Figure 7 seem to exhibit some periodicity; we suspect that this is somehow related to the fact that our algorithm depends on partitioning a rectangle and therefore some pattern persists among the various instances of C. The individual approximation ratios for each of the 50 states are shown in the Appendix.

6 Conclusions

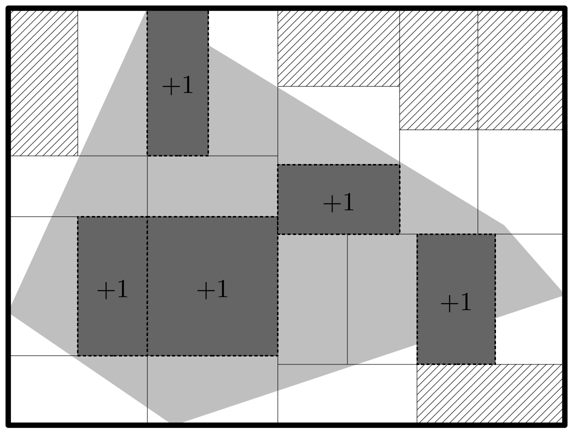
We have presented a fast and simple approximation algorithm for choosing the k medians of the continuum of demand points defined by a convex polygon C. Although the theoretical worst-case approximation ratio of our algorithm is 2.74 using our current estimates of upper and lower bounds, we find that in practice the algorithm generally finds solutions within 10% of optimality.

7 Acknowledgments

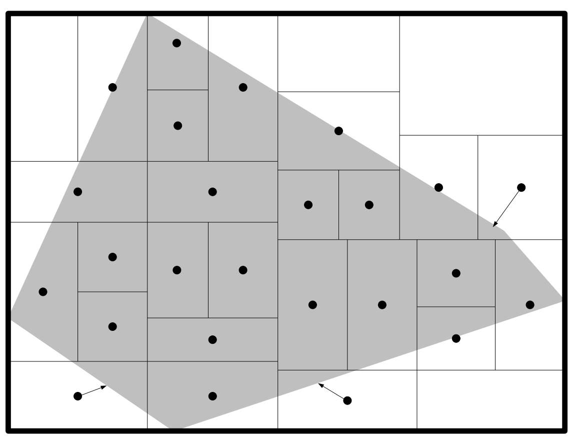
The authors thank Joseph Mitchell, Joseph O'Rourke, Stan Wagon, and an anonymous referee for their helpful suggestions.

References

[1] B. Aronov, P. Carmi, and M.J. Katz. Minimum-cost load-balancing partitions. *Algorithmica*, 54(3):318–336, July 2009.



(a) The hatched rectangles indicate the members of `OUT`, where we have $k = 25$; since $|\texttt{OUT}| = 5$, we have 5 extra points to distribute among the members of `IN`, which are all the unhatched rectangles (note that one member of `IN` is not a final rectangle). The "+1" symbols correspond to rectangles for which $\tilde{N}_j = N_j + 1$ (ties are broken arbitrarily).



(b) The output of our modified algorithm; the small arrows indicate that points located outside C are translated towards C accordingly.

Figure 6: The effect of our minor algorithmic modification.

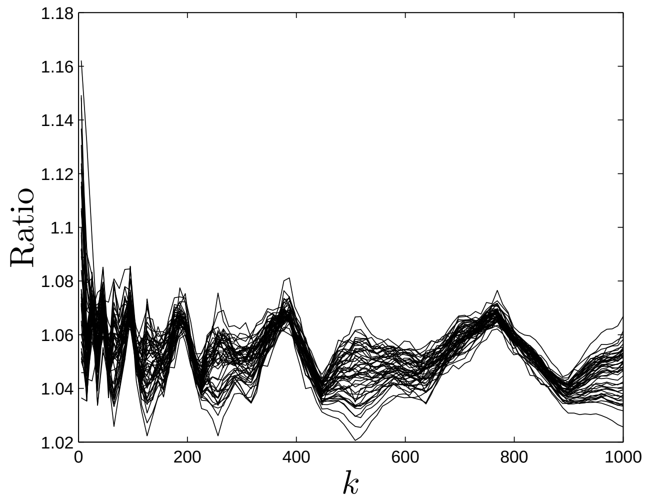


Figure 7: The approximation ratios for $k \leq 1000$ applied to the 50 states of the USA.

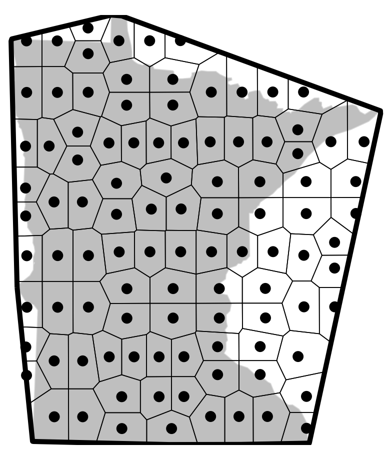


Figure 8: The output of our modified algorithm (and the induced Voronoi diagram) for $k = 100$ when applied to the convex hull of the state of Minnesota.

[2] Sanjeev Arora, Prabhakar Raghavan, and Satish Rao. Approximation schemes for Euclidean k-medians and related problems. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, STOC '98, pages 106–113, New York, NY, USA, 1998. ACM.

[3] Jack Brimberg, Pierre Hansen, Nenad Mlandinovic, and Eric. D. Taillard. Improvement and comparison of heuristics for solving the uncapacitated multisource weber problem. *Oper. Res.*, 48:444–460, May 2000.

[4] Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem. *J. Comput. Syst. Sci.*, 65:129–149, August 2002.

[5] Sándor P. Fekete, Joseph S. B. Mitchell, and Karin Beurer. On the continuous Fermat-Weber problem. *Oper. Res.*, 53:61–76, January 2005.

[6] J. N. Lyness and R. Cools. A survey of numerical cubature over triangles. In *Proceedings of Symposia in Applied Mathematics*, pages 127–150. American Mathematical Society, 1994.

[7] Joseph S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems. *SIAM J. Comput.*, 28:1298–1309, March 1999.

[8] C. H. Papadimitriou. Worst-case and probabilistic analysis of a geometric location problem. *SIAM Journal on Computing*, 10:542, 1981.

[9] Franco P. Preparata and Michael I. Shamos. *Computational geometry: an introduction*. Springer-Verlag New York, Inc., New York, NY, USA, 1985.

[10] G. Toussaint. Solving geometric problems with the rotating calipers. *Proc IEEE Melecon*, 9(May):1–8, 1983.

Appendix

Tables 1 through 5 show the approximation ratios for $k \leq 1000$. We also show the Voronoi diagrams and point placements for $k = 25$ for additional clarity.

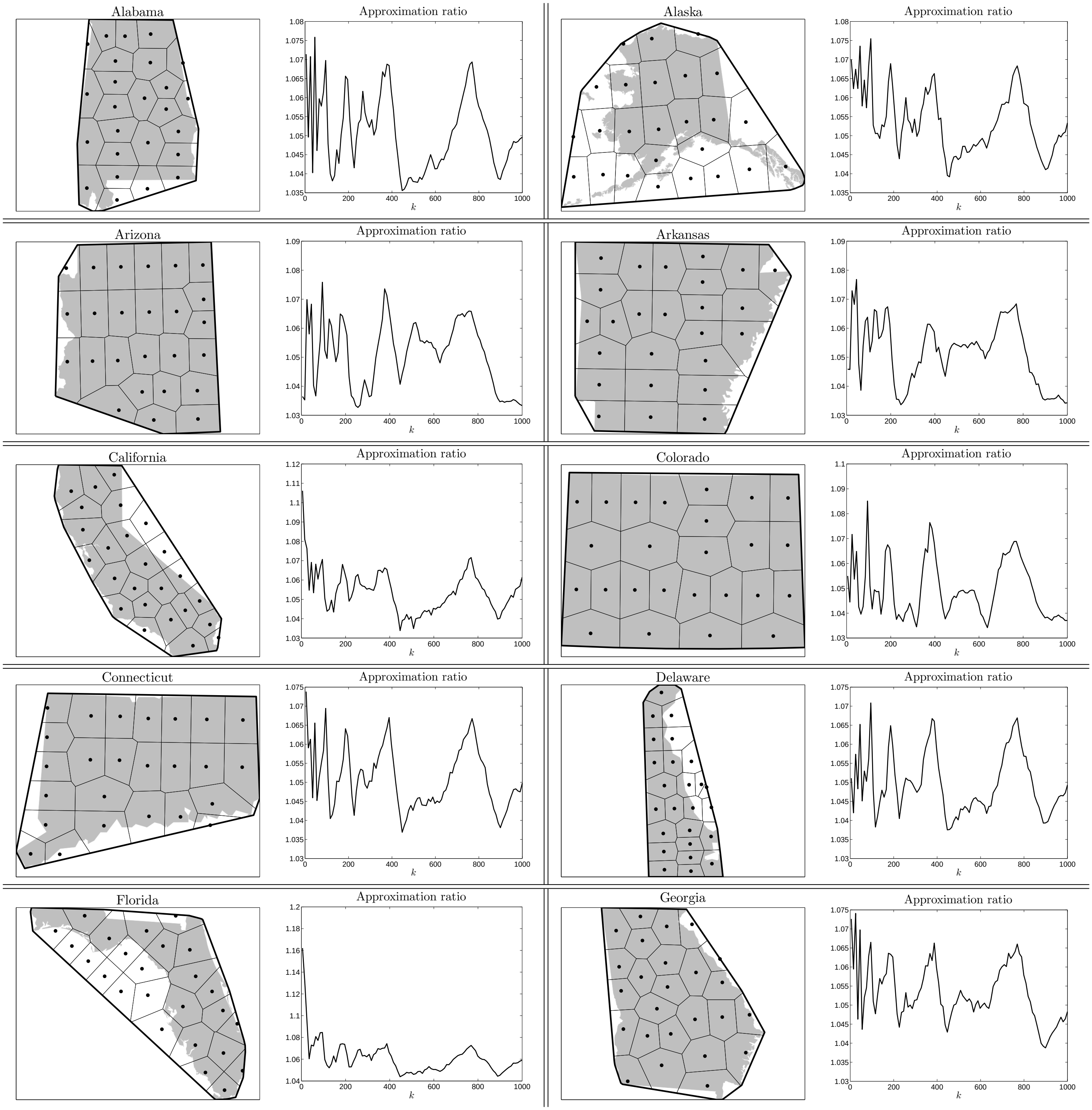


Table 1: Approximation ratios and point configurations for $k = 25$ for Alabama through Georgia.

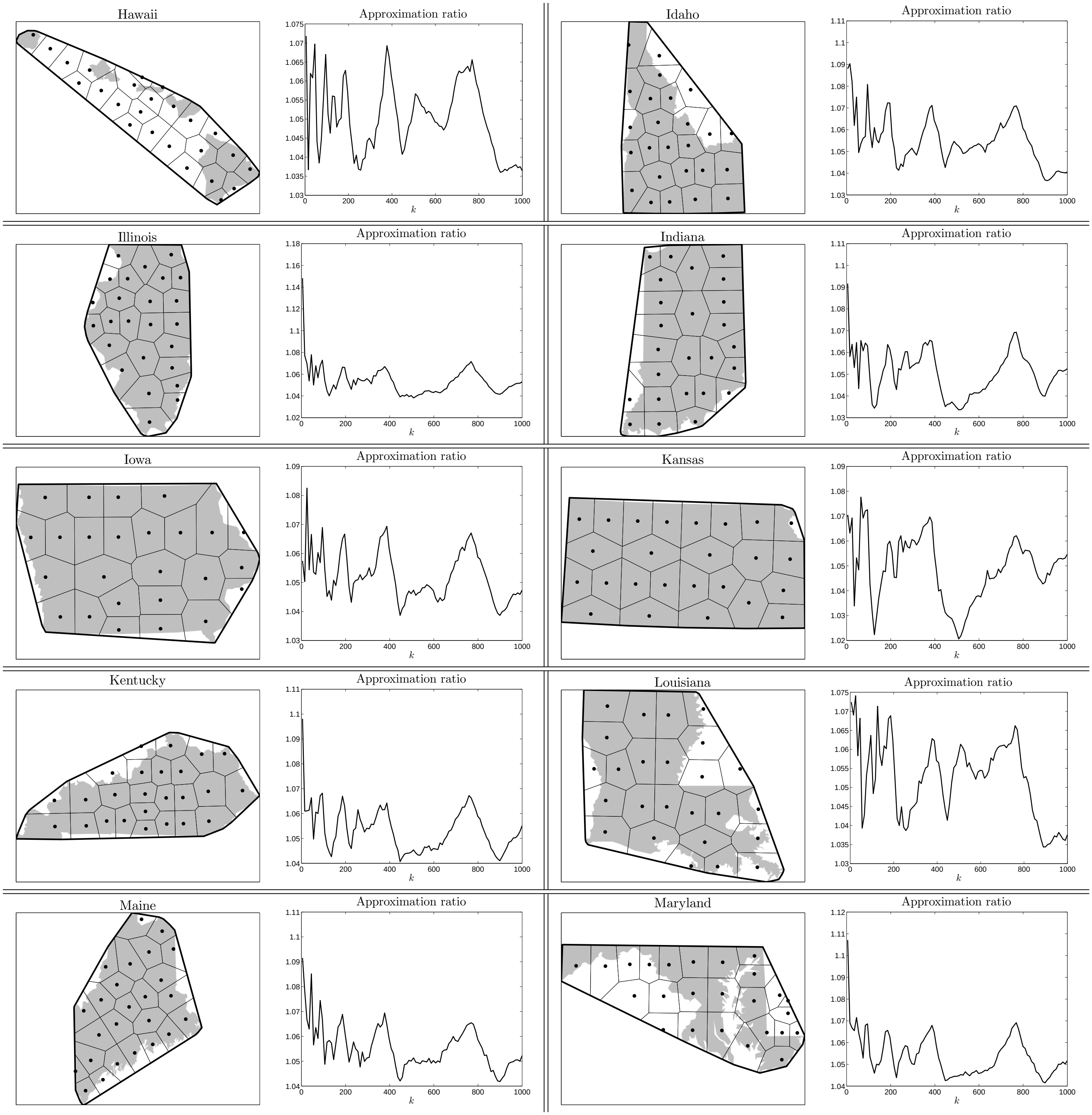


Table 2: Approximation ratios and point configurations for $k = 25$ for Hawaii through Maryland.

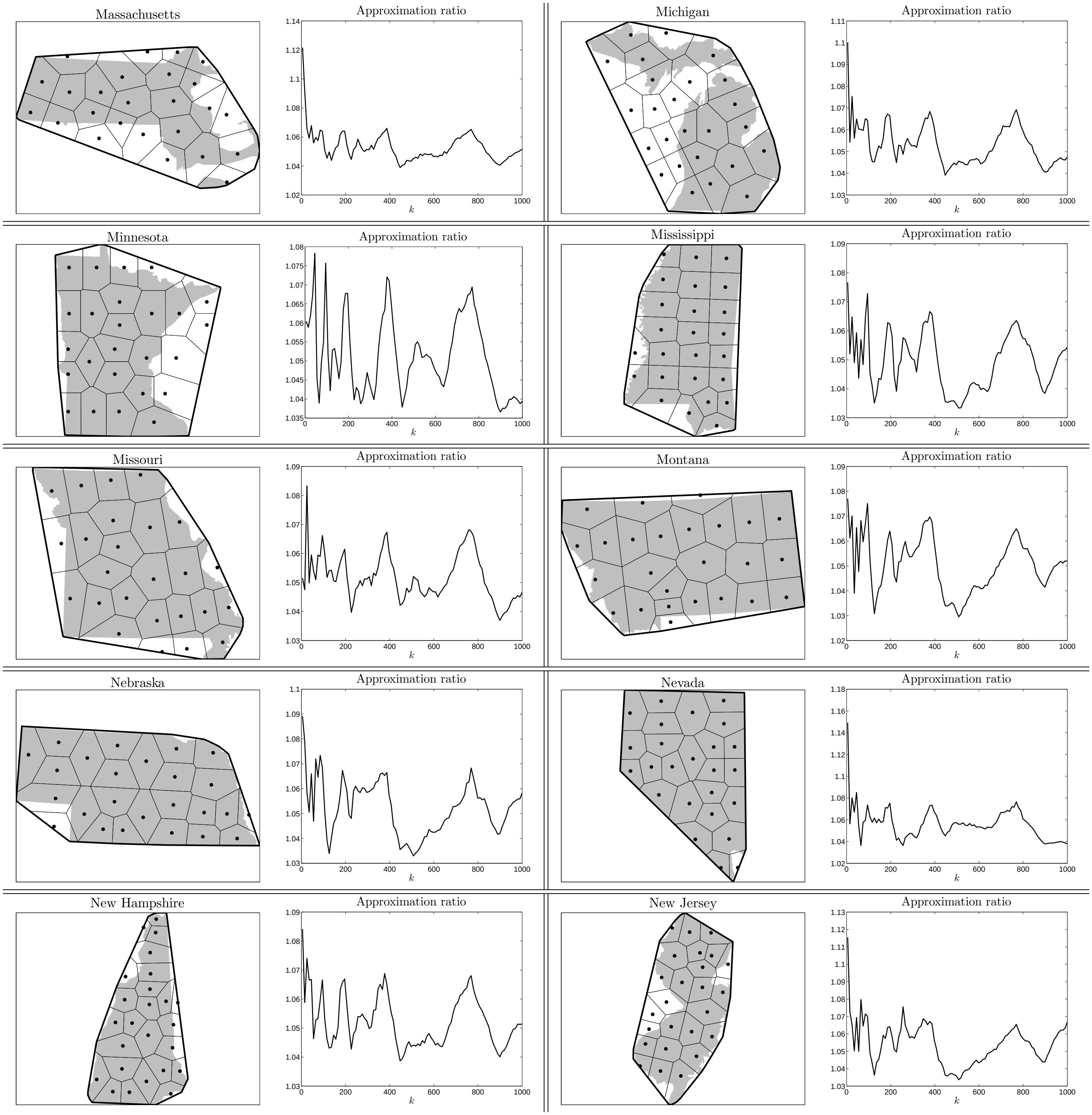


Table 3: Approximation ratios and point configurations for $k = 25$ for Massachusetts through New Jersey.

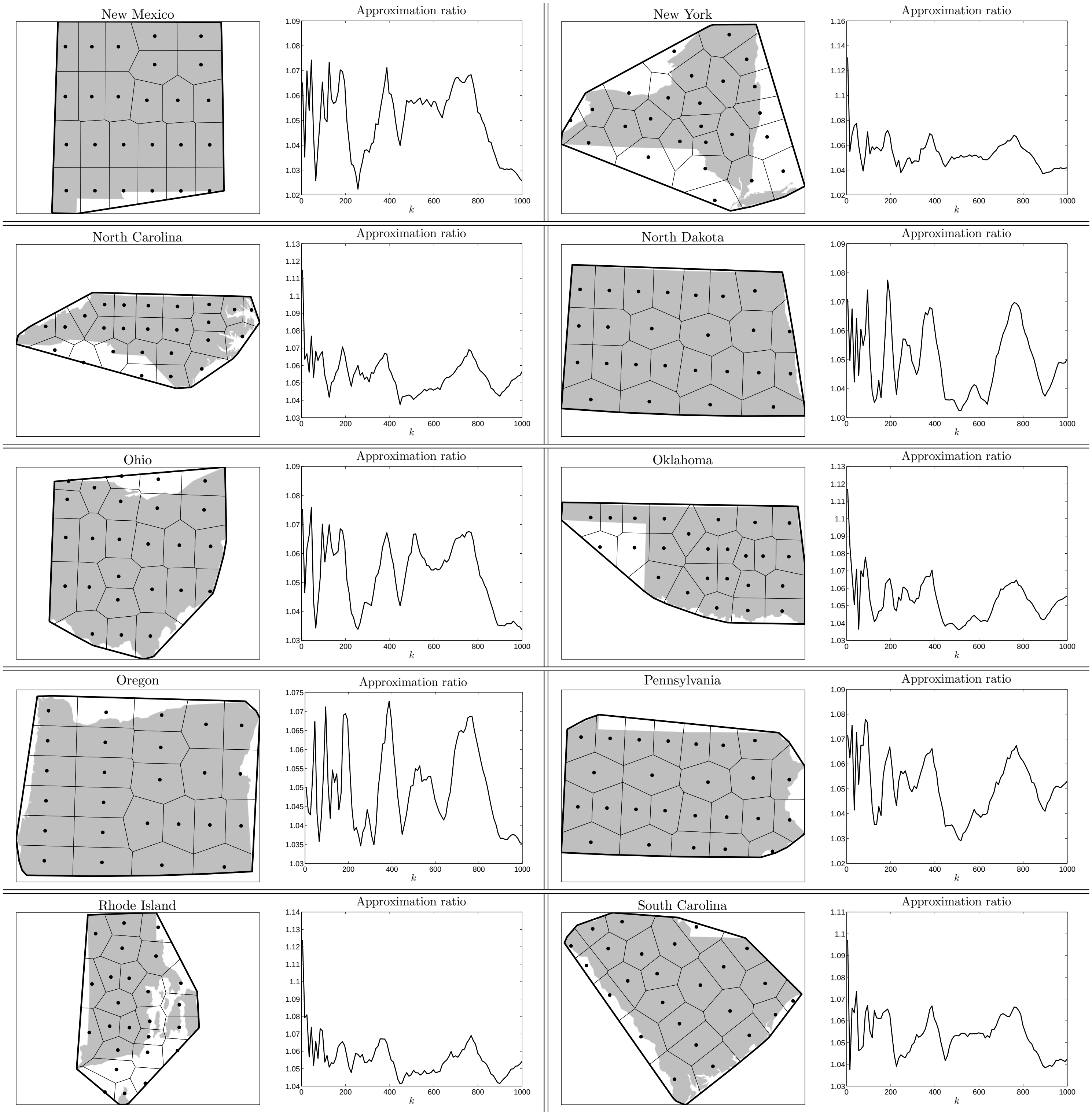


Table 4: Approximation ratios and point configurations for $k = 25$ for New Mexico through South Carolina.

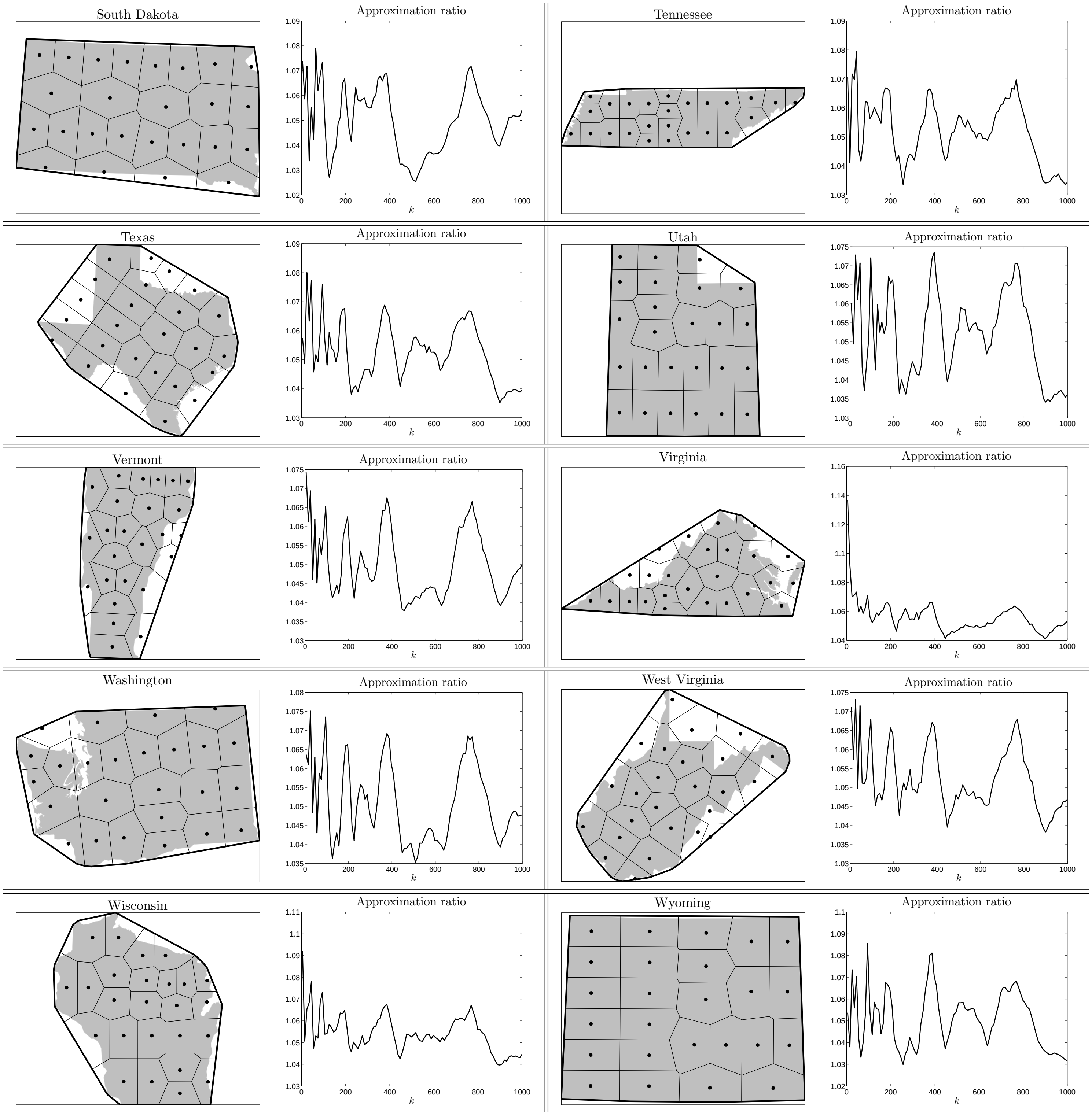


Table 5: Approximation ratios and point configurations for $k = 25$ for South Dakota through Wyoming.