Discrete Applied Mathematics 2 (1980) 295–309

PROBABILISTIC ANALYSIS OF SOME EUCLIDEAN CLUSTERING PROBLEMS

A.M. FRIEZE

Dept. of Computer Science and Statistics, Queen Mary College, Mile End Road, London, E1 4NS, England

Received 28 May 1979
Revised 2 July 1980

We are given n points distributed randomly in a compact region D of R^m. We consider various optimisation problems associated with partitioning this set of points into k subsets. For each problem we demonstrate lower bounds which are satisfied with high probability. For the case where D is a hypercube we use a partitioning technique to give deterministic upper bounds and to construct algorithms which with high probability can be made arbitrarily accurate in polynomial time for a given required accuracy.

1. Introduction

We are given n points $X=\{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}\}$ belonging to a given compact region $D\subseteq R^m$. We study in this paper various optimisation problems associated with such a set:

Problem 1. Find $Y=\{\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(k)}\}\subseteq D$ such that

$$z_1(X,Y)=\max(\min(\|\mathbf{x}^{(i)}-\mathbf{y}^{(j)}\|:j=1,\ldots,k):i=1,\ldots,n)$$

is minimised.

Problem 2. Find $Y=\{\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(k)}\}\subseteq X$ such that $z_1(X,Y)$ is minimised. It will be convenient to refer to the objective function as $z_2(X,Y)$ in this case.

Problem 3. Partition X into k subsets $X_1,\ldots,X_k$ so that

$$z_3(X_1,\ldots,X_k)=\max(\max(\|\mathbf{x}-\mathbf{y}\|:\mathbf{x},\mathbf{y}\in X_j):j=1,\ldots,k)$$

is minimised.

Problem 4. Partition X into k subsets $X_1,\ldots,X_k$ so that

$$z_4(X_1,\ldots,X_k)=\max\Bigl(\sum_{\mathbf{x},\mathbf{y}\in X_j}\|x-y\|:j=1,\ldots,k\Bigr)$$

is minimised.

The norms considered will be

$$\|\boldsymbol{x}\|_e = \left(\sum_{j=1}^{m} x_j^2\right)^{1/2},$$

$$\|x\|_\infty = \max(|x_j| : j = 1, \dots, m).$$

Non-euclidean versions of the above problems are known to be NP-hard as are the corresponding problems of finding ε-optimal solutions for arbitrary $\varepsilon > 0$.

(For $m = 1$ problems 1, 2, 3 are solvable in polynomial time using dynamic programming. the status of problem 4 when $m = 1$ is not known.)

It is likely therefore that problems 1–4 are also NP-hard as is the case for Euclidean versions of other NP-hard problems [2, 3]. This paper conducts a probabilistic analysis of these problems. The n points are assumed to be randomly and uniformly distributed over the region D which is assumed to have hyper-volume V.

Results can be obtained for other norms by using the fact that for any two norms $\|\ \|_a$, $\|\ \|_b$ there exists a constant p such that for $\boldsymbol{x} \in R^m$ $\|x\|_a \leqslant p\,\|\boldsymbol{x}\|_b$. For example if $m = 2$ and k, n grow so that $k/n \to 0$ as $n \to \infty$ we show that in problem 1 using $\|\ \|_e$ that

$$z_1^* = \min z_1(X, Y) \geqslant (V/k\pi)^{1/2}$$

with probability tending to 1. Now as $\|\boldsymbol{x}\|_\infty \geqslant \|\boldsymbol{x}\|_e/\sqrt{2}$ this implies that using $\|\ \|_\infty$

$$z_1^* \geqslant (V/2k\pi)^{1/2}$$

with probability tending to 1. We can however prove in this case that $z_1^* \geqslant \frac{1}{2}(V/k)^{1/2}$ with probability tending to 1. We have thus analysed these norms separately.

We follow the approach used in Fisher and Hochbaum [1]. For an instance of problem t we denote the value of an optimal solution by $z_t^*(n, k)$. For each problem we derive lower bounds for z_t^* which are valid with probability tending to 1 assuming that $k/n \to d < 1$ in problems, 1, 2, 3 and $d \leqslant \frac{1}{2}$ for problem 4.

Then restricting our attention to the case where D is a hypercube we derive simple upper bounds for z_t^*. We then use a grid technique as in Fisher and Hochbaum [1] such that given $\varepsilon > 0$ we derive a solution of value $\hat{z}_t$ where $\hat{z}_t - z_t^* \leqslant \varepsilon z_t^*$ with probability tending to 1. The time complexity of these algorithms are $O(n^{p(\varepsilon)})$ where $p(\varepsilon)$ naturally depends on ε. Fisher and Hochbaum analysed the k-median problem: find $Y = \{\mathbf{y}^{(j)}, \dots, \mathbf{y}^{(k)}\} \subseteq X$ such that

$$\sum_{j=1}^{n} \min(\|\boldsymbol{x}^{(j)} - \mathbf{y}^{(i)}\|_e : i = 1, \dots, k)$$

is minimised.

They only considered $m = 2$ and $\|\ \|_e$ but their analysis would extend easily to general m.

The results obtained here can be usefully compared with those of [1], most importantly for problem 2 with $m=2$ and $\| \|_e$ we show that for a fixed region the optimal value (usually) grows like $1/\sqrt{k}$ whereas for the k-median problem the optimal value grows like $n/\sqrt{k}$. The factor n is what one would expect on comparing objective functions.

2. Analysis of problem 1

We first compute a probabilistic lower bound to problem 1 using $\| \|_e$. We shall use Stirling's inequalities

$$(n/e)^n(2n\pi)^{1/2} \leqslant n! \leqslant (12n/12n-1)(n/e)^n(2n\pi)^{1/2}$$

several times to replace factorials and so we have stated them here for convenience.

Notation. For $a \in R$, $a \geqslant 0$ and $\mathbf{c} \in R^m$ the hypersphere is

$$\mathrm{HS}(\mathbf{c}, a) = \{x \in R^m : \|\mathbf{x}-\mathbf{c}\|_e \leqslant a\}.$$

It's hypervolume is denoted by $c_m a^m$ where the c_m satisfy

$$c_1 = 2 \quad \text{and} \quad c_{m+1} = \left(2\int_0^{\pi/2} \cos^{m+1}\theta \, \mathrm{d}\theta\right)c_m \quad \text{for } m \geqslant 1.$$

Note that

$$\int_0^{\pi/2} \cos^{2n}\theta \, \mathrm{d}\theta = \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdots \frac{1}{2} \cdot \frac{\pi}{2},$$

$$\int_0^{\pi/2} \cos^{2n+1}\theta \, \mathrm{d}\theta = \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdots \frac{2}{3}.$$

Let $X \subseteq R^m$ be finite. Let $r = r(X)$ be the radius of the smallest hypersphere containing X and let $\mathbf{c} = \mathbf{c}(X)$ be the centre of this hypersphere.

Lemma 2.1. *Let X, r, $\mathbf{c}$ be as above and suppose $r \leqslant a$. If $\mathbf{z} \in R^m$ is such that $r\{X \cup \{\mathbf{z}\}) \leqslant a$, then $\|\mathbf{z}-\mathbf{c}\|_e \leqslant a + (a^2 - r^2)^{1/2}$.*

Proof. Let $Y = \{\mathbf{x} \in X : \|\mathbf{x}-\mathbf{c}\|_e = r\} \neq \emptyset$. Now let C = convex hull of Y. We show by contradiction that $\mathbf{c} \in C$. If $\mathbf{c} \notin C$ let $\boldsymbol{b}$ be the nearest point of C to $\mathbf{c}$. Let $\mathbf{c}_\lambda = (1-\lambda)\mathbf{c} + \lambda\boldsymbol{b}$ for $0<\lambda<1$. Now for $\lambda>0$ and $\mathbf{y} \in Y$ $\|\mathbf{y}-\mathbf{c}_\lambda\|_e < \|\mathbf{y}-\mathbf{c}\|_e$ and so if λ is "small enough" $\mathbf{c}$ can be the centre of a hypersphere of radius $r' < r$ containing X. Thus $\mathbf{c} \in C$ and so $c = \sum_{i=0}^d \lambda_i \mathbf{y}_i$ where $\lambda_i > 0$ for $i = 1, \ldots, d$ and $\sum_{i=1}^d \lambda_i = 1$. Now let $\mathbf{z}$ be as in the statement of the lemma and let $\mathbf{c}_1 = \mathbf{c}(X \cup \{z\})$.

Since $c \in C$ there exists $\mathbf{y}_t$ such that $(\mathbf{c}_1 - \mathbf{c}) \cdot (\mathbf{y}_t - \mathbf{c}) \leqslant 0$. Then

$$\begin{aligned} a^2 &\geqslant \|\mathbf{c}_1 - \mathbf{y}_t\|_e^2 \\ &= \|\mathbf{c}_1 - \mathbf{c}\|_e^2 + \|\mathbf{y}_t - \mathbf{c}\|_e^2 - 2(\mathbf{c}_1 - \mathbf{c}) \cdot (\mathbf{y}_t - \mathbf{c}) \\ &\geqslant \|\mathbf{c}_1 - \mathbf{c}\|_e^2 + r^2. \end{aligned}$$

Thus $\|\mathbf{c}_1 - \mathbf{c}\|_e \leqslant (a^2 - r^2)^{1/2}$ and hence

$$\|z - \mathbf{c}\|_e \leqslant \|z - \mathbf{c}_1\|_e + \|\mathbf{c}_1 - \mathbf{c}\|_e \leqslant a + (a^2 - r^2)^{1/2}. \quad \square$$

Lemma 2.2. *Let $E(n, \mathbf{c}, a)$ be the event that n points $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ chosen at random in D lie in* HS$(\mathbf{c}, a)$ *and let $F(n, a) = \bigcup_{\mathbf{c} \in D} (n, \mathbf{c}, a)$.*

Then for $n \geqslant 2$ there exists $b = b(m) > 1$ such that

$$P(n, a) = \text{Prob}(F(n, a)) \leqslant b^{\sqrt{n}} v^{n-1} \tag{2.1}$$

where $v = c_m a^m / V$.

Proof. Let $p(n, z)$ be the density function of the random variable z = radius of the smallest hypersphere containing n random points in D.

It follows that $P(n, a) = \int_0^a p(n, z)\,\mathrm{d}z$ and it follows from Lemma 2.1 that

$$P(n+1, a) \leqslant (c_m/V) \int_0^a p(n, z)(a + (a^2 - z^2)^{1/2})^m \,\mathrm{d}z.$$

Integration by parts gives

$$P(n+1, a) \leqslant (c_m a^m / V) P(n, a) + (mc_m/V) \int_0^a P(n, z)(a + (a^2 - z^2)^{1/2})^{m-1} z (a^2 - z^2)^{-1/2} \,\mathrm{d}z. \tag{2.2}$$

Now $P(2, a) \leqslant 2^m c_m a^m / V$ and if for some $n \geqslant 2$ and constant α $P(n, z) \leqslant \alpha (c_m z^m / V)^{n-1}$ for all $z \geqslant 0$ then substitution in (2.2) gives

$$P(n+1, a) \leqslant \alpha (1 + u_{n-1})(c_m a^m / V)^n$$

where

$$u_{n-1} = m \int_0^{\pi/2} (1 + \cos\theta)^{m-1} (\sin\theta)^{m(n-1)+1} \,\mathrm{d}\theta.$$

We deduce therefore that for $n \geqslant 2$

$$P(n, a) \leqslant 2^m \prod_{t=1}^{n-2} (1 + u_t)(c_m a^m / V)^{n-1}. \tag{2.3}$$

The RHS of (2.3) is bounded by $b^{\sqrt{n}} v^{n-1}$ for some b dependent on m. This can be shown as follows:

$$u_n \leq m 2^{m-1} \int_0^{\pi/2} (\sin\theta)^{mn+1}\,\mathrm{d}\theta = \alpha \frac{mn}{mn+1}\cdot\frac{mn-2}{mn-1}\cdots$$

where α is dependent on m. If

$$\beta_M = \frac{M}{M+1}\cdot\frac{M-2}{M-1}\cdots\frac{2}{3} \quad \text{for } M \text{ even,}$$

$$= \frac{M}{M+1}\cdot\frac{M-2}{M+1}\cdots\frac{1}{2} \quad \text{for } M \text{ odd,}$$

we show that $\beta_M < 2/\sqrt{M}$. For M even

$$\beta_M < 1\cdot\frac{M-1}{M}\cdot\frac{M-3}{M-2}\cdots\frac{3}{4}.$$

Thus $\beta_M^2 < 2/(M+1) < 2/M$. For M odd $\beta_M^2 < 1/M$ by a similar argument. Thus

$$u_n < 2\alpha/\sqrt{mn} = \beta/\sqrt{n} \quad \text{for } \beta = 2\alpha/\sqrt{m}.$$

Thus for $n \geq 2$

$$\prod_{t=1}^{n-2}(1+u_t) < \prod_{t=1}^{n-2}(1+\beta/\sqrt{t}) < \mathrm{e}^{2\beta\sqrt{n-2}}$$

as may be shown by induction on n.

Thus the R.H.S. of (2.3) $< 2^m \mathrm{e}^{2\beta\sqrt{n-2}} v^{n-1}$ which can be simplified to $b^{\sqrt{n}} v^{n-1}$ for large enough b.

Lemma 2.3. *Let n points $X = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ be chosen at random in D. For $a \geq 0$ and $v = c_m a^m / V$*

$$\operatorname{Prob}(z_1^*(n,k) \leq a) \leq (kv)^{n-k} \mathrm{e}^k b^{\sqrt{nk}} / \sqrt{2\pi k} \tag{2.4}$$

if $\|\ \|_e$ is used.

Proof. Let $X_1, \ldots, X_k \in \operatorname{PART}(n,k)$ = the set of (unordered) partitions of X into k subsets. Let $n_t = |X_t|$ for $t = 1, \ldots, k$ and let

$$Q = \operatorname{Prob}((X_t \subseteq \operatorname{HS}(\mathbf{c}_t, a) \text{ for some } \mathbf{c}_t \in D) \text{ for } t = 1, \ldots, k)$$

$$= \prod_{t=1}^{k} \operatorname{Prob}(X_t \subseteq \operatorname{HS}(\mathbf{c}_t, a) \text{ for some } \mathbf{c}_t \in D)$$

$$= \prod_{t=1}^{k} \operatorname{Prob}(F(n_t, a)) \quad \text{as in Lemma 2.2.}$$

By Lemma 2.2

$$Q \leq v^{n-k} \prod_{t=1}^{k} b^{\sqrt{n_t}} \leq v^{n-k} b^{\sqrt{nk}}.$$

Now let $S(X)=\{(X_1,\ldots,X_k)\in \mathrm{PART}(n,k): r(X_t)\leqslant a$ for $t=1,\ldots,k\}$ (r as in Lemma 2.1). We note that $S(X)=\emptyset \rightarrow z_1^*(n,k)>a$. Thus

$$\begin{aligned}\mathrm{Prob}(z_1^*(n,k)\leqslant a)&\leqslant \mathrm{Prob}(S(X)\neq\emptyset)\\ &\leqslant E(|S(X)|)\quad\text{(Expectation by the above)}\\ &\leqslant |\mathrm{PART}(n,k)|\, v^{n-k}b^{\sqrt{nk}}\\ &\leqslant (k^n/k!)v^{n-k}b^{\sqrt{nk}}\end{aligned}$$

The result now follows after using Stirlings inequalities. □

Theorem 2.1. *For sequence of problems where $n\rightarrow\infty$ and $k=nd+\mathrm{O}(1/n^3)$ with $0\leqslant d<1$ we have*

$$\mathrm{Prob}(z_1^*(n,k)\leqslant(\alpha_d V/kc_m)^{1/m}\leqslant 1/\sqrt{2\pi k}+\mathrm{O}(1/n) \tag{2.5}$$

where $\alpha_d=(\mathrm{e}^{-d}b^{-\sqrt{d}})^{1/(1-d)}$.

Proof. Simply substitute $(\alpha_d V/kc_m)^{1/m}$ for a in (2.4). □

Thus if $k\rightarrow\infty$ (2.5) provides a lower bound for $z_1^*(n,k)$ with probability tending to 1.

For constant k we must clearly have $z_1^*\geqslant(V/kc_m)^{1/m}$ – note $\alpha_0=1$ – else we cannot cover D with k hyperspheres of radius z_1^*. It is straightforward to show that for finite k we must do this with probability tending to 1.

We continue by computing lower bounds to z_1^* for $\|\ \|_\infty$. We base our analysis on a lemma about covering D with hypercubes. It will be used for sets of the form $\{\boldsymbol{x}\in R^m:\|\boldsymbol{x}-\mathbf{c}\|\leqslant a\}$.

Notation. Let $\boldsymbol{a},\mathbf{c}\in R^m\boldsymbol{a}\geqslant 0$. The hyperoblong is

$$\mathrm{HO}(\mathbf{c},\boldsymbol{a})=\{\boldsymbol{x}\in R^m:|x_j-c_j|\leqslant\tfrac{1}{2}a_j \text{ for } j=1,\ldots,m\}.$$

It's hypervolume is of course $a_1a_2\cdots a_m$.

Lemma 2.5. *Let $E(n,\mathbf{c},\boldsymbol{a})$ be the event that n points $X=\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(n)}\}$ chosen at random from D lie in $\mathrm{HO}(\mathbf{c},\boldsymbol{a})$ and let $F(n,a)=\bigcup_{\mathbf{c}\in D}E(n,\mathbf{c},\boldsymbol{a})$. Then for $n\geqslant 2$*

$$P(n,\boldsymbol{a})=\mathrm{Prob}(F(n,\boldsymbol{a}))\leqslant n^m v^{n-1}$$

where $v=a_1a_2\cdots a_m/V$.

Proof. Let $p(n,\boldsymbol{z})$ be the density function of the random vector $\boldsymbol{z}\in R^m$ where $z_1,\ldots,z_m$ are the lengths of the sides of the smallest hyperoblong containing the set X. These lengths are given by

$$z_i=\max(x_i^{(1)},\ldots,x_i^{(n)})-\min(x_i^{(1)},\ldots,x_i^{(n)}).$$

Thus

$$P(n, \boldsymbol{a}) = \int_0^{a_1} \cdots \int_0^{a_m} p(n, \boldsymbol{z})\, \mathrm{d}z_m \cdots \mathrm{d}z_1 \quad \text{and} \quad p(n, z) = \frac{\partial^m p(n, z)}{\partial z_1 \cdots z_m}.$$

We also have

$$P(n+1, \boldsymbol{a}) \leqslant \int_0^{a_1} \cdots \int_0^{a_m} (p(n, \boldsymbol{z})/V) \Big(\prod_{i=1}^{m} (2a_i - z_i) \Big)\, \mathrm{d}z_m \cdots \mathrm{d}z_1. \tag{2.6}$$

This is because for given $z_1 \cdots z_m$ the random point $\boldsymbol{x}^{(n+1)}$ must lie in a hyperoblong of sides $(2a_1 - z_1) \cdots (2a_m - z_m)$ in order that $F(n+1, \boldsymbol{a})$ can occur.

Next let $M = \{1, 2, \ldots, m\}$ and for $S \subseteq M$ let P_S denote $P(n, h_1 \cdots h_m)$ where $h_i = a_i$ for $i \in S$ and $h_i = z_i$ for $i \notin S$. Let $\mathrm{d}_S = \prod_{i \notin S} \mathrm{d}z_i$ and $a_S = \prod_{i \in S} a_i$. Successive integration of the RHS of (2.6) by parts gives

$$P(n+1, \boldsymbol{a}) \leqslant \Big(\sum_{S \subseteq M} a_S \int P_S \, \mathrm{d}_S \Big) \Big/ V. \tag{2.7}$$

Now $P(2, \boldsymbol{a}) \leqslant 2^m a_1 \cdots a_m / V$ and if for some $n \geqslant 2$ and constant α $P(n, z) \leqslant \alpha((\prod_{i=1}^m z_i)/V)^{n-1}$ for $\boldsymbol{z} \geqslant 0$ then from (2.7) we have

$$\begin{aligned} P(n+1, a) &\leqslant \Big(\sum_{s \subseteq M} \alpha a_S^n \int \Big(\prod_{i \notin S} z_i^{n-1} \, \mathrm{d}z_i \Big) \Big/ V^n \\ &= \Big(\sum_{S \subseteq M} \alpha a_M^n n^{-|\bar{S}|} \Big) \Big/ V^n, \quad \bar{S} = M - S \\ &= \alpha (1 + 1/n)^m v^n. \end{aligned}$$

Thus

$$P(n+1, \boldsymbol{a}) \leqslant \prod_{t=1}^{n} (1 + 1/t)^m v^n = (n+1)^m v^n. \quad \square$$

Lemma 2.7. *Let n points $X = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\}$ be chosen at random in D. For $a \geqslant 0$ and $v = a^m / V$*

$$\mathrm{Prob}(z_1^*(n, k) \leqslant \tfrac{1}{2} a) \leqslant (kv)^{n-k} \mathrm{e}^k (n/k)^{km} / \sqrt{2\pi k}. \tag{2.8}$$

Proof. Let $(X_1, \ldots, X_k) \in \mathrm{PART}(n, k)$ and $|X_i| = n_i$ for $i = 1, \ldots, k$. Let $\boldsymbol{a} = (a, \ldots, a) \in R^m$ and let $Q = \mathrm{Prob}((X_i \subseteq \mathrm{HO}(\boldsymbol{c}_i, \boldsymbol{a})$ for some $\boldsymbol{c}_i \in D)$ for $i = 1, \ldots, k$.

By Lemma 2.4 with $v = a^m / V$

$$Q \leqslant \prod_{i=1}^{k} n_i^m v^{n_i - 1} \leqslant (n/k)^{km} v^{n-k}.$$

It follows as in Lemma 2.3 that

$$\text{Prob}(z_1^*(n,k)\leq a/2)\leq (k^n/k!)Q.$$

The result now follows after using Stirlings inequalities. □

Theorem 2.2. *For a sequence of problems where* $n\to\infty$ *and* $k=nd+\mathrm{O}(1/n)$ *with* $0\leq d<1$ *we have*

$$\text{Prob}(z_1^*(n,k)\leq (\tfrac{1}{2}\alpha_d V/k)^{1/m})\leq 1/\sqrt{2\pi k}+\mathrm{O}(1/n) \tag{2.9}$$

where $\alpha_d=(\mathrm{e}^{-d}d^{md})^{1/(1-d)}$.

Proof. Simply substitute $(\alpha_d V/k)^{1/m}$ for a in (2.8). □

Similar comments to those given after Theorem 2.1 apply. We now describe the calculation of upper bounds and approximate solutions in the case that D is a hypercube of side L. Let $\hat{k}=\lfloor k^{1/m}\rfloor$ and divide D uniformly into $\hat{k}^m$ hypercubes of side $L/\hat{k}$ and let Y consist of the centres of these hypercubes plus $k-\hat{k}^m$ other points in D.

$\|\ \|_e$. For $\boldsymbol{x}\in D$ there is a point $\mathbf{y}\in Y$ such that $\|\boldsymbol{x}-\mathbf{y}\|_e\leq m^{1/2}L/2\hat{k}$ and so for this norm

$$z_1^*(n,k)\leq m^{1/2}L/2\hat{k}.$$

$\|\ \|_\infty$. For $\boldsymbol{x}\in D$ there is a point $\mathbf{y}\in Y$ such that $\|\boldsymbol{x}-\mathbf{y}\|_\infty\leq L/2\hat{k}$ and so for this norm

$$z_1^*(n,k)\leq L/2\hat{k}.$$

Notice that if $k=\hat{k}^m$ this upper bound coincides closely with the lower bound derived after Theorem 2.2 when $d=0$.

We now consider approximate solutions. Let $t>0$ be an integer which determines the proposed accuracy of the solution. Divide D uniformly into $T=t^m$ hypercubes $H_1,\ldots,H_T$ of side L/t. Let $C=\{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_T\}$ be the set of centres of these hypercubes. Let

$$\hat{z}_1=\min(z_1(X,Y):Y\subseteq C \text{ and } |Y|=k).$$

This can be computed in $\mathrm{O}(2^T nk)$ time. Let Y^* minimise z_1. Assume without loss of generality that $Y^*\subseteq\bigcup_{j=1}^k H_j$. Now for $\boldsymbol{x}\in D$ and $\mathbf{y}\in H_j$

$$\|\boldsymbol{c}_j-\boldsymbol{x}\|\leq\|\boldsymbol{c}_j-\mathbf{y}\|+\|\mathbf{y}-\boldsymbol{x}\|$$

and hence

$$\hat{z}_1-z_1^*\leq\max(\|\boldsymbol{c}_1-\mathbf{y}\|:\mathbf{y}\in Y^*\cap H_1).$$

$\|\ \|_e$. Thus $\hat{z}_1-z_1^*\leq m^{1/2}L/2t$. Now fix $1>\varepsilon>0$ and consider a sequence of problems for which $k\leq p\log n$ where $p>0$. Putting $t=\lceil m^{1/2}(kc_m)^{1/m}/2\varepsilon\rceil$ we see that $\hat{z}_1-z_1^*\leq\varepsilon z_1^*$ with probability $\geq 1-(2\pi k)^{-1/2}$.

For large k $2^T \approx A^k$ where $A = 2^{(m^{m/2}c_m/2^m\varepsilon^m)}$

$\leqslant n^{p \log A}$

and so the approximation scheme is polynomial when k is restricted in this manner.

$\| \|_\infty$. In this case $\hat{z}_1 - z_1^* \leqslant L/2t$ and we take $t = \lceil k^{1/m}/\varepsilon \rceil$.

3. Analysis of problem 2

We first compute a probabilistic lower bound for problem 2 using $\| \|_e$.

Lemma 3.1. *Let n points* $X = \{x^{(1)}, \ldots, x^{(n)}\}$ *be chosen at random in D. For* $a \geqslant 0$ *and* $v = c_m a^m / V$

$$\operatorname{Prob}(z_2^*(n, k) \leqslant a) \leqslant (12/11)(kv)^{n-k}(n^n/k^k(n-k)^{n-k})\sqrt{n/2\pi k(n-k)}. \tag{3.1}$$

Proof. Let $J = \{j_1, \ldots, j_k\} \subseteq N = \{1, 2, \ldots, n\}$ and let $Y = \{x^{(j_1)}, \ldots, x^{(j_k)}\}$. If $j \in N - J$, then Prob(there exists $i(j) \in J$ such that $\|x^{(j)} - x^{(i(j))}\|_e \leqslant a\} \leqslant kv$. Hence

$$\text{Prob(for all } j \in N-J \text{ there exists } i(j) \in J \text{ such that } \|x^{(j)} - x^{(i(j))}\|_e \leqslant a) \leqslant (kv)^{n-k} \tag{3.2}$$

Now there are $\binom{n}{k}$ subsets of size k in N and hence $\operatorname{Prob}(z_2^*(n, k) \leqslant a) = \operatorname{Prob}((3.2)$ holds for some $J) \leqslant \binom{n}{k}(kv)^{n-k}$. the result now follows after using Stirlings inequalities.

Theorem 3.1. *For a sequence of problems where* $n \to \infty$ *and* $k = nd + O(1/n)$ *with* $0 \leqslant d < 1$ *we have*

$$\operatorname{Prob}(z_2^*(n, k) \leqslant (\alpha_d V/kc_m)^{1/m}) \leqslant (12/11)\sqrt{n/2\pi k(n-k)} + O(1/n) \tag{3.3}$$

where $\alpha_d = (1-d)d^{d/(1-d)}$.

Proof. Use Lemma 3.1. □

Thus if $k \to \infty$ (3.3) provides a lower bound for $z_2^*(n, k)$ with probability tending to 1.

For constant k the problem can be solved exactly in $O(n^{k+1})$ time by examining each k-subset of X.

In the case of $\| \|_\infty$ a similar proof gives

Theorem 3.2. *For a sequence of problems where* $n \to \infty$ *and* $k = nd + O(1/n)$ *with* $0 \leqslant d < 1$ *we have*

$$\operatorname{Prob}(z_2^*(n, k) \leqslant \tfrac{1}{2}(\alpha_d V/k)^{1/m}) \leqslant (12/11)\sqrt{n/2\pi k(n-k)} + O(1/n) \tag{3.3}$$

where $\alpha_d = (1-d)d^{d/(1-d)}$.

We once again describe the calculation of upper bounds and approximate solutions in the case that D is a hypercube of side L. We again divide D uniformly into $\hat{k}^m$ hypercubes of side $L/\hat{k}$ and this time to produce Y we select one point of X from each hypercube that contains points of X and then make up Y to size k be arbitrary addition of points in X not used so far. This gives

$$z_2^* \leqslant m^{1/2} L/\hat{k} \quad \text{for } \|\ \|_e,$$
$$z_2^* \leqslant L/\hat{k} \quad \text{for } \|\ \|_\infty.$$

To obtain approximate solutions we proceed in much the same manner as in Section 2. We choose $t > 0$ as before and divide D into $H_1, \ldots, H_T$. For each $J \subseteq SJ = \{J \subseteq \{1, \ldots, T\} : |J| = k\}$ we proceed as follows: for each $j \in J$ such that $H_j \cap X \neq \emptyset$ choose $\boldsymbol{x}^{(j)} \in H_j \cap X$. This produces $k_1 \leqslant k$ points to which we arbitrarily add $k - k_1$ other points from X to form a set $Y(J)$. Then let $\hat{z}_2 = \min(z_2(X, Y(J)) : J \subseteq SJ)$ which can be computed in $O(2^T nk)$ time.

Now let Y^* minimize z_2 and assume without loss of generality that $Y^* \subseteq \bigcup_{j=1}^k H_j$. A use of the triangular inequality as in Section 2 shows that

$$\hat{z}_2 - z_2^* \leqslant L_t \quad \text{where } L_t = \max(\|\boldsymbol{x} - \boldsymbol{y}\| : \boldsymbol{x}, \boldsymbol{y} \in H_1).$$

Assuming $k \leqslant d \log n$ and given $\varepsilon > 0$ and taking

$$t = \lceil m^{1/2} (kc_m)^{1/m} / \varepsilon \rceil \quad \text{for } \|\ \|_e,$$
$$t = \lceil 2k^{1/m} / \varepsilon \rceil \quad \text{for } \|\ \|_\infty, \tag{3.4}$$

we have $\hat{z}_2 - z_2^* \leqslant \varepsilon z_2^*$ with high probability and the time taken is polynomial in n.

4. Analysis of problem 3

Our lower bounds for $\|\ \|_e$ are based on

Lemma 4.1. *Let $X \subseteq R^m$ be a finite set and suppose that $\boldsymbol{x}, \boldsymbol{y} \in X$ implies $\|\boldsymbol{x} - \boldsymbol{y}\|_e \leqslant a$. Then $r = r(X) \leqslant a(m/2(m+1))^{1/2}$ where r is the radius of the smallest hypersphere containing X.*

Proof. Let $\boldsymbol{c} = \boldsymbol{c}(X)$ be the centre of this hypersphere and as in Lemma 2.1 $\boldsymbol{c} = \sum_{i=1}^d \lambda_i \boldsymbol{y}_i$ where $\|\boldsymbol{y}_i - \boldsymbol{c}\|_e = r$. We can assume by Caratheodory's theorem that $d \leqslant m+1$. If $\boldsymbol{z}_i = (\boldsymbol{y}_i - \boldsymbol{c})/r$ for $1 \leqslant i \leqslant d$ then

$$0 = \left\| \sum_{i=1}^d \lambda_i \boldsymbol{z}_i \right\|_e^2 = \sum \lambda_i^2 + 2 \sum \lambda_i \lambda_j \boldsymbol{z}_i \cdot \boldsymbol{z}_j$$

We show that there exists k, l such that $\boldsymbol{z}_k \cdot \boldsymbol{z}_l \leqslant -1/(d-1)$. (If $d = 1$ then $X = \{\boldsymbol{c}\}$ and the result is trivial.) For if not we have

$$0 > \sum \lambda_i^2 - (2/(d-1)) \sum \lambda_i \lambda_j = \left(\sum (\lambda_i - \lambda_j)^2\right)/(d-1) \geqslant 0.$$

Thus

$$\begin{aligned} a^2 &\geq \|\mathbf{y}_k - \mathbf{y}_l\|_e^2 \\ &= r^2 \|\mathbf{z}_k - \mathbf{z}_l\|_e^2 \\ &= r^2(\mathbf{z}_k^2 + \mathbf{z}_l^2 - 2\mathbf{z}_k \cdot \mathbf{z}_l) \\ &\geq r^2(2 + 2/(d-1)) \geq r^2(2 + 2/m). \quad \square \end{aligned}$$

Using this result in conjunction with Theorem 2.1 gives

Theorem 4.1. *For a sequence of problems where* $n \to \infty$ *and* $k = nd + O(1/n)$ *with* $0 \leq d < 1$ *we have*

$$\text{Prob}(z_3^*(n, k) \leq \alpha_d (V/kc_m)^{1/m}) \leq 1/\sqrt{2\pi k} + O(1/n) \tag{4.1}$$

where $\alpha_d = (2(m+1)/m)^{1/2}(e^{-d} b^{-\sqrt{d}})^{1/m(1-d)}$.

The result for $\| \, \|_\infty$ depends on the fact that if $X \subseteq R^m$ is such that $\mathbf{x}, \mathbf{y} \in X$ implies $\|\mathbf{x} - \mathbf{y}\|_\infty \leq a$ then X can be contained in a hypercube of side a. This gives using Theorem 2.2.

Theorem 4.2. *For a sequence of problems for which* $n \to \infty$ *and* $k/n \to d < 1$ *we have*

$$\text{Prob}(z_3^*(n, k) \leq (\alpha_d V/k)^{1/m}) \leq 1/\sqrt{2\pi k} + O(1/n) \tag{4.2}$$

where $\alpha_d = (e^{-d} d^{md})^{1/(1-d)}$.

Once again assuming that D is a hypercube of side L we obtain upper bounds by dividing D into $\bar{k} = \hat{k}^m$ hypercubes of side $L/\hat{k}$. Let these hypercubes be $H_1, \ldots, H_{\bar{k}}$. We then partition X into $X \cap H_1, \ldots, X \cap H_{\bar{k}}$ plus $k - \bar{k}$ empty sets. If there are points of X on the boundaries of several hypercubes we assign these points arbitrarily to one of them. This partition gives

$$\begin{aligned} z_3^* &\leq m^{1/2} L/\hat{k} \quad \text{for } \| \, \|_e \\ z_3^* &\leq L/\hat{k} \qquad \text{for } \| \, \|_\infty. \end{aligned}$$

To obtain approximate solutions we again choose $t > 0$ and divide D into $T = t^m$ hypercubes $H_1, \ldots, H_T$. For $J \subseteq \{1, 2, \ldots, T\}$ let $H_J = \bigcup_{i \in J} H_i$ and let $P(T) =$ the set of partitions of $\{1, \ldots, T\}$ into k subsets. For $(J_1, \ldots, J_k) \in P(T)$ let

$$Z_3(J_1, \ldots, J_k) = \max(\max(\|\mathbf{x} - \mathbf{y}\| : \mathbf{x}, \mathbf{y} \in H_{J_i} \cap X) : i = 1, \ldots, k)$$

and let $\hat{z}_3 = \min(Z_3(J_1, \ldots, J_k) : (J_1, \ldots, J_k) \in P(T))$. $\hat{z}_3$ can be computed in $O((k^T/k!)n^2)$ time. Now let $(X_1^*, \ldots, X_k^*)$ be the optimal partition for z_3. The partitions generated in computing $\hat{z}_3$ are all those that satisfy

$$X_i \cap H_r \neq \emptyset \text{ for some } i, r \text{ implies } X_j \cap H_r = \emptyset \text{ for } i \neq j. \tag{4.3}$$

If $X_1^*, \ldots, X_k^*$ does not satisfy (4.3) then we can find $(\hat{X}_1, \ldots, \hat{X}_k)$ satisfying (4.3) and

$$z_3(\hat{X}_1, \ldots, \hat{X}_k) \leq z_3(X_1^*, \ldots, X_k^*) + 2L_t. \tag{4.4}$$

$(\hat{X}_1, \ldots, \hat{X}_k)$ is obtained by starting with $(X_1^*, \ldots, X_k^*)$ and while there are $r, i, j_1, \ldots, j_p$ contravening (4.3) amending the current partition $(X_1, \ldots, X_k)$ by $X_i := X_i \cup \bigcup_{s=1}^{p} (X_{j_s} \cap H_r)$ and $X_{j_s} := X_{j_s} - H_r$ for $s = 1, \ldots, p$. We observe that throughout the above process

$$\boldsymbol{x} \in X_i \text{ implies there exists } \mathbf{y}, r \text{ such that } y \in X_i^* \text{ and } \boldsymbol{x}, \mathbf{y} \in H_r \tag{4.5}$$

(either $\mathbf{y} = \boldsymbol{x}$ or prior to some change of partition $\boldsymbol{x} \in X_{j_i} \cap H_r$ and it is then moved to X_i while $\mathbf{y} \in X_i \cap H_r$ is never moved).

Let $z_3(\hat{X}_1, \ldots, \hat{X}_k) = \|\hat{\boldsymbol{x}} - \hat{\mathbf{y}}\|$ where $\hat{\boldsymbol{x}}, \hat{\mathbf{y}} \in \hat{X}_p$. From (4.5) there exist $\boldsymbol{x}^*, \mathbf{y}^*, r, s$ such that $\boldsymbol{x}^*, \mathbf{y}^* \in X_p^*$, $\hat{\boldsymbol{x}}, \boldsymbol{x}^* \in H_r$ and $\hat{\mathbf{y}}, \mathbf{y}^* \in H_s$. Thus

$$\begin{aligned} \|\hat{\boldsymbol{x}} - \hat{\mathbf{y}}\| &\leq \|\hat{\boldsymbol{x}} - \boldsymbol{x}^*\| + \|\boldsymbol{x}^* - \mathbf{y}^*\| + \|y^* - \hat{\mathbf{y}}\| \\ &\leq L_t + z_3^* + L_t \end{aligned}$$

and thus $\hat{z}_3 \leq z_3^* + 2L_t$.

$\|\ \|_e$. Here $\hat{z}_3 - z_3^* \leq 2m^{1/2}L/t$. Now fix $1 < \varepsilon < 0$ and take

$$t = \lceil m(kc_m)^{1/m}/((m+1)/2)^{1/2}\varepsilon \rceil.$$

This gives $z_3 - z_3^* \leq \varepsilon z_3^*$ with high probability. The dominant term in $k^T/k!$ is, using Stirling's formulae A^k for some constant A and for polynomial time we again assume $k \leq d \log n$.

$\|\ \|_\infty$. Here we take $t = \lceil 2k^{1/m}/\varepsilon \rceil$.

5. Analysis of problem 4

We first compute a probabilistic lower bound to problem 4 using $\|\ \|_e$.

Lemma 5.1. *Let* $X = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\}$ *be chosen at random in* D. *Let* $z(X) = \sum_i \sum_j \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\|_e$. *Then*

$$\text{Prob}(z(X) \leq a) \leq n(m!\, c_m (a/n)^m / V)^{n-1} / (m(n-1))!. \tag{5.1}$$

Proof. We first consider the following: $\boldsymbol{c}$ is an arbitrary point of D and $Y = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(q)}\}$ are chosen at random in D. Let $d_i = \|\boldsymbol{c} - \mathbf{y}^{(i)}\|_e$ and $d(\mathbf{c}, Y) = \sum_i d_i$. Then

$$\text{Prob}(b_i \leq d_i \leq b_i + \delta b_i) \leq mc_m b_i^{m-1}\, \delta b_i (1 + \text{O}(\delta b_i))/V$$

as $mc_m b_i^{m-1}\, \delta b_i$ is the approximate hypervolume of a "thin hyperannulus" of radius b_i and thickness δb_i. Consequently

$$\text{Prob}(d(c,Y)\leq a)\leq \int_{b_1=0}^{b_1=a}\int_{b_2=0}^{b_2=a-b_1}\cdots\int_{b_q=0}^{b_q=a-\sum_1^{q-1}b_i}\prod_i (mc_m b_i^{m-1}\,\mathrm{d}b_i/V)$$
$$=(m!\,c_m a^m/V)^q/(qm)!$$

which is easily proved by induction. Putting $X_j=X/\{\boldsymbol{x}^{(j)}\}$ we see that

$$\text{Prob}(d(\boldsymbol{x}^{(j)},X_j)\leq a)\leq (m!\,c_m a^m/V)^{n-1}/((n-1)m)!. \tag{5.2}$$

Now

$$\text{Prob}\left(\sum_{j=1}^{n} d(\boldsymbol{x}^{(j)},X_j)\leq na\right)\leq \text{Prob}(d(\boldsymbol{x}^{(j)},X_j)\leq a \text{ for at least one } j=1,\ldots,n)$$
$$\leq n\,\text{Prob}(d(\boldsymbol{x}^{(1)},X_1)\leq a). \tag{5.3}$$

We obtain (5.1) by replacing a in (5.3) by a/n and using (5.2). □

Lemma 5.2. *Let $X=\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(n)}\}$ be chosen at random in D. For $a\geq 0$ and assuming $n\geq 2k$, then*

$$\text{Prob}(z_4^*(n,k)\leq a)\leq A/B \tag{5.4}$$

where

$$A=\mathrm{e}^{m(n-1)+k}k^{(2m+1)n-2(m+1)k}n^k(m!\,c_m a^m/V)^{n-k},$$
$$B=(m(n-k)^2)^{m(n-k)}(2\pi mk(n-k))^{1/2}.$$

Proof. Let $(X_1,\ldots,X_k)\in\text{PART}(n,k)$ and let $n_t=|X_t|$ for $t=1,\ldots,k$ and let $Q=\text{Prob}(z(X_t)\leq a$ for $t=1,\ldots,k)$ where z is as defined in the statement of Lemma 4.1. From this lemma

$$Q\leq\prod_{t=1}^{k} n_t(m!\,c_m a^m/V)^{n_t-1}/n_t^{m(n_t-1)}(m(n_t-1))!$$
$$\leq (n/k)^k(m!\,c_m a^m/V)^{n-k}/P$$

where

$$P=\prod_{n_t\geq 2}(mn_t(n_t-1))^{m(n_t-1)}(2\pi m(n_t-1))^{1/2}\mathrm{e}^{-m(n_t-1)}$$
$$\geq m^{m(n-k)}((n-k)/k)^{2m(n-k)}(2\pi m(n-k))^{1/2}\mathrm{e}^{-m(n-1)}$$

where $n\geq 2k$ is used in one of the reductions.

As there are at most $k^n/k!$ partitions we have our result in the usual way after using Stirling's inequalities. □

Theorem 5.1. *For a sequence of problems where $n\to\infty$ and $k=nd+\mathrm{O}(1/n)$, and*

(i) $0<d\leqslant\frac{1}{2}$:

$$\text{Prob}(z_4^*(n,k)\leqslant\alpha_d(V/kc_m)^{1/m})\leqslant 1/\sqrt{2\pi mk(n-k)}+\text{O}(1/n)$$

where $\alpha_d=((1-d)/d)^2m((d^d\text{e}^{-m-d})^{1/(1-d)}/m!)^{1/m}$.

(ii) $d=0$;

$$\text{Prob}(z_4^*(n,k)\leqslant\beta_m(n/k)^2n^{-k/n}(V/kc_m)^{1/m})\leqslant 1/\sqrt{2\pi mk(n-k)}+\text{O}(1/n)$$

where $\beta_m=m/\text{e}(m!)^{1/m}$.

Proof. Use Lemma 5.2. □

For $\|\ \|_\infty$ we have

Theorem 5.2. *For a sequence of problems where* $n\to\infty$ *and* $k=nd+\text{O}(1/n)$ *and*

(i) $0<d\leqslant\frac{1}{2}$:

$$\text{Prob}(z_4^*(n,k)\leqslant\alpha_d(V/k)^{1/m}/2)\leqslant 1/\sqrt{2\pi mk(n-k)}+\text{O}(1/n).$$

(ii) $d=0$:

$$\text{Prob}(z_4^*(n,k)\leqslant\beta_m(n/k)^2n^{-k/n}(V/k)^{1/m}/2)\leqslant 1/\sqrt{2\pi mk(n-k)}+\text{O}(1/n)$$

where α_d, β_m *are as in Theorem 5.1.*

Once again assuming that D is a hypercube of side L we obtain an upper bound by using the partition defined in Section 4. This gives

$$z_4^*\leqslant m^{1/2}Ln^2/\hat{k}\quad\text{for }\|\ \|_e,$$

$$z_4^*\leqslant Ln^2/\hat{k}\qquad\text{for }\|\ \|_\infty.$$

Note that these upper bounds are larger than the lower bounds by a factor of order of magnitude k^2. This can be explained by the possibility that all the points of X lie in the same small hypercube. To obtain approximate solutions we proceed as in Section 4 to compute

$$\hat{z}_4=\min(Z_4(J_1,\ldots,J_k):(J_1,\ldots,J_k)\in P(T))$$

where

$$Z_4(J_1,\ldots,J_k)=\max\Bigg(\sum_{\boldsymbol{x},\boldsymbol{y}\in H_{J_i}\cap X}\|\boldsymbol{x}-\boldsymbol{y}\|:i=1,\ldots,k\Bigg)$$

assuming the usual uniform division of D into t^m hypercubes. Once again let $(X_1^*,\ldots,X_k^*)$ minimise z_4. If this partition satisfies (4.3) then $\hat{z}_4=z_4^*$ otherwise we can compute from it a partition $(\hat{X}_1,\ldots,\hat{X}_k)$ satisfying (4.3) and

$$z_4(\hat{X}_1,\ldots,\hat{X}_k)\leqslant z_4^*+n^2L_t. \tag{5.5}$$

We start from $(X_1^*,\ldots,X_k^*)$ and a general stage of the construction suppose we have the partition $(X_1,\ldots,X_k)$ and for some r $|I|\geqslant 2$ where $I=\{i:X_i\cap H_r\neq\emptyset\}$. Let $\mathbf{c}$ be the centre of H and for $i\in I$ let $d_i=\sum_{\boldsymbol{x}\in X_i}\|\boldsymbol{x}-\mathbf{c}\|$ and let $d_p=\min(d_i)$. We amend the current partition as follows:

$$X_p := X_p \cup (H_r \cap X)),$$
$$X_i := X_i - H_r, \quad i \in I - \{p\}.$$

The change Δ in the value of z_4 is

$$\sum_{i \in I-\{p\}} \sum_{x \in H_r \cap X_i} \left(\sum_{y \in X_p} \|x - y\| - \sum_{y \in X_i - H_r} \|x - y\| \right).$$

Now

$$\sum_{x \in H_r \cap X_i} \sum_{y \in X_p} \|x - y\| \leqslant n_i (d_p + |X_p|\, L_t/2)$$

where $n_i = |H_r \cap X_i|$ and

$$\sum_{x \in H_r \cap X_i} \sum_{y \in X_i - H_r} \|x - y\| \geqslant n_i (d_i - |X_i|\, L_t/2)$$

and so

$$\Delta \leqslant \sum_{i \in I - \{p\}} n_i (d_p - d_i + (|X_p| + |X_i|) L_t/2)$$
$$\leqslant \sum_{i \in I - \{p\}} n_i n L_t.$$

Continuing this process until (4.3) is satisfied we find that the total change is $\leqslant n^2 L_t$.

$\| \ \|_e$. Here $\hat{z}_4 - z_4^* \leqslant 2m^{1/2} n^2 L/t$. Now fix $1 < \varepsilon < 0$ and take

$$t = \lceil 2m^{1/2} k^2 (kC_m)^{1/m} / \varepsilon \beta m \rceil.$$

This gives $\hat{z}_4 - z_4^* \leqslant \varepsilon z_4^*$ with high probability. The time for computing z_4 is $O(n^2 k^{T+1}/k!)$ and the dominant term in $k^T/k!$ is one of $k^{k^{2+1/m}}$ and to get a polynomial time algorithm we assume $k \leqslant (b \log n / \log \log n)^{1/(2+1/m)}$.

$\| \ \|_\infty$. Here we take $t = \lceil 2k^{2+1/m}/\varepsilon \beta m \rceil$.

Acknowledgement

I am grateful to Colin McDiarmid for his valuable comments.

References

[1] M.L. Fisher and D.S. Hochbaum, Probabilistic analysis of the euclidean k-median problem, Research Report 78-06-03, University of Pennsylvania (1978).

[2] M.R. Garey, R.L. Graham and D.S. Johnson, Some NP-complete geometric problems, Proceedings 8th ACM Symposium on Theory of Computing (1976) 10–29.

[3] C.H. Papadimitriou, The euclidean travelling salesman problem is NP-complete, Theoretical Computer Science 4 (1977) 237–244.