ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE

G. Rinaldi, U. Voigt, G.J. Woeginger

THE MATHEMATICS OF PLAYING GOLF, OR: A NEW CLASS OF DIFFICULT NON-LINEAR MIXED INTEGER PROGRAMS

R. 553 Settembre 2001

Giovanni Rinaldi – Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy – email: rinaldi@iasi.rm.cnr.it.

Ulrich Voigt – Institut für Mathematik, Universität Freiburg, D-79104 Freiburg im Breisgau, Germany – email: uvo@mathematik.uni-freiburg.de.

Gerhard J. Woeginger – Department of Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, and Institut für Mathematik, Technische Universität Graz, Steyrergasse 30, A-8010 Graz, Austria – email: g.j.woeginger@math.utwente.nl.

Research carried out with financial support of the project TMR-DONET nr. ERB FMRX–CT98–0202 of the European Union. The third author was supported by the START program Y43-MAT of the Austrian Ministry of Science.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica, CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.rm.cnr.it
URL: http://www.iasi.rm.cnr.it

Abstract

We consider a class of non-linear mixed integer programs with n integer variables and k continuous variables. Solving instances from this class to optimality is an NP-hard problem. We show that for the cases with $k = 1$ and $k = 2$, every optimal solution is integral. In contrast to this, for every $k \geq 3$ there exist instances where every optimal solution takes non-integral values.

Keywords: non-linear optimization – mixed integer program – integrality – computational complexity – NP-hard problem – golf problem.

1. Introduction

In February 2001, the following mathematical puzzle lead to a long discussion in the newsgroup `rec.denksport.de`:

> *A golf course consists of nine holes with distances of 150, 225, 250, 275, 300, 325, 350, 400, and 425 meters. A golfer only knows how to perform two different strokes; one of them brings the ball to some distance x, and the other stroke brings the ball to a distance y. Every stroke must be done in a straight line towards the hole. However a stroke may go beyond the hole, and then the subsequent strokes must be done back towards the hole. How should x and y be chosen such that the whole golf course can be mastered with the minimum possible number of strokes?*

Setting $x = 75$ and $y = 175$ yields a solution with 26 strokes: $150 = x + x$, $225 = x + x + x$, $250 = x+y$, $275 = y+y-x$, $300 = x+x+x+x$, $325 = x+x+y$, $350 = y+y$, $400 = x+x+x+y$, $425 = x + y + y$. Another solution with 26 strokes results from setting $x = 125$ and $y = 150$. Are 26 strokes the best possible solution, or is there also a better solution that uses 25 strokes, or even less? And how does one show that some solution is best possible? All nine distances are integer multiples of 25. Does this imply that in any optimal solution x and y must be integer multiples of 25? These and several related questions will be answered in this paper.

In a more general and more mathematical formulation of this problem, the input consists of the positive integers $d_1, \ldots, d_n$ together with a positive integer k. The goal is to find k stroke lengths $s_1, \ldots, s_k$ such that the n distances $d_1, \ldots, d_n$ can be represented with the minimum number of strokes:

$$\begin{array}{lll} \min & \sum_{i=1}^{n} \sum_{j=1}^{k} |z_{ij}| & \\ \text{s.t.} & \sum_{j=1}^{k} s_j \cdot z_{ij} = d_i & \text{for } i = 1, \ldots, n \\ & s_j \text{ non-negative real} & \text{for } j = 1, \ldots, k \\ & z_{ij} \text{ integer} & \text{for } i = 1, \ldots, n,\ j = 1, \ldots, k \end{array}$$

Here z_{ij} denotes the number of strokes of length s_j that are performed for the hole with distance d_i. If z_{ij} is positive, then z_{ij} strokes are made towards the hole; if z_{ij} is negative, then $|z_{ij}|$ strokes are made back towards the hole. This problem is called the *golf problem.* The restricted special case of the golf problem where the number k of stroke lengths is a fixed constant is called the *k-golf problem.* We use the term *optimal integral solution* to denote the best solution of the golf problem subject to the additional restriction that $s_1, \ldots, s_k$ are integers.

In this paper, we derive several combinatorial and algorithmical results on the golf problem. In Section 2 there are some preliminary results and observations: It is shown that the golf problem is contained in the complexity class NP, and that the k-golf problem is solvable in pseudo-polynomial time for every fixed k. Section 3 shows that for $k = 2$, every optimal solution must have integral stroke lengths. Section 4 shows that for every $k \geq 3$, there exist instances for which every optimal solution must have fractional stroke lengths. Section 5 proves NP-hardness of the golf problem, and Section 6 gives some conclusions and open questions.

2. First observations on the golf problem

We use $\gcd(d_1, \ldots, d_n)$ to denote the greatest common divisor of the n integers $d_1, \ldots, d_n$. For an instance of the golf problems with distances $d_1, \ldots, d_n$, we denote $D = \sum_{i=1}^{n} d_i$ and

$d_{\max} = \max_{1 \le i \le n} d_i$. Observe that the sizes of D and $d_{\max}$ (i.e., the number of bits needed to represent these numbers) are polynomially bounded in the size of the n integers $d_1, \ldots, d_n$.

The golf problem for $k = 1$ is straightforward to solve.

Lemma 2.1. *For an instance of the golf problem with distances $d_1, \ldots, d_n$, the optimal solution for $k = 1$ stroke lengths uses the stroke length $s_1 = \gcd(d_1, \ldots, d_n)$. The optimal number of strokes equals $D/s_1 = \sum_{i=1}^{n} d_i / \gcd(d_1, \ldots, d_n)$.*

Lemma 2.2. *For any instance of the golf problem with distances $d_1, \ldots, d_n$, the optimal solution and the optimal integral solution use at most D strokes.*

Proof. A trivial feasible solution uses only the stroke length $s = 1$ and makes D strokes. ∎

Lemma 2.3. *For any instance of the golf problem with distances $d_1, \ldots, d_n$ and k stroke lengths, there exists an optimal solution in which all stroke lengths are rational numbers whose numerators and denominators are bounded by D^{2k^2}.*

Proof. By Lemma 2.2, in any optimal solution all numbers z_{ij} satisfy $|z_{ij}| \le D$. We fix all values z_{ij} in an arbitrary optimal solution, but let the values s_j float. We choose an index set $K \subseteq \{1, 2, \ldots, n\}$ such that the equations $\sum_{j=1}^{k} s_j \cdot z_{ij} = d_i$ with $i \in K$ form an independent subsystem of maximal rank. Clearly, $|K| \le k$. Then the values s_j, $j = 1, \ldots, k$, may be fixed as a basic feasible solution of an $k \times |K|$ system of equations where all coefficients and all values in the right hand sides are bounded by D; if the basic feasible solution contains a negative s_t, then we flip the signs of the corresponding coefficients z_{it}. The statement of the lemma follows now from linear algebra folklore (see e.g., Schrijver [3]). ∎

Lemma 2.4. *The decision version of the golf problem is contained in the complexity class NP.*

Proof. By Lemmas 2.2 and 2.3, every instance of the golf problem has an optimal solution where all z_{ij} have their size bounded by $O(\log_2 D)$ and all s_j have their size bounded by $O(k^2 \log_2 D)$. Hence, there exists a certificate with size polynomially bounded in the instance size. ∎

Lemma 2.5. *For any fixed value k, the k-golf problem is solvable in pseudo-polynomial time.*

Proof. By Lemma 2.3 there exists a solution where all stroke lengths are rational numbers whose numerators and denominators are bounded by D^{2k^2}. Since k is a fixed constant, this yields a pseudo-polynomial number $O(D^{4k^3})$ of possible cases for the values of the k numerators and the k denominators. We will separately consider every such case with fixed stroke lengths. Once the values s_j are fixed, the golf problem boils down to solving n instances of a k-dimensional integer program, one instance for every distance d_i, $i = 1, \ldots, n$. Since integer programming in fixed dimension is polynomially solvable (Lenstra [2]), every single case can be solved in polynomial time. ∎

3. The case with two stroke lengths

In this section we consider the case of two stroke lengths. Suppose that there exists an instance for $k = 2$ with a non-integral optimal solution s_1 and s_2. By Lemma 2.3 we may assume that s_1 and s_2 are rational, and without loss of generality $s_1 > s_2$. Let N be the least common denominator of s_1 and s_2 such that $s_1 = x/N$ and $s_2 = y/N$ for two positive integers x and y. The equations $z_{i1}s_1 + z_{i2}s_2 = d_i$ imply that

$$z_{i1}x + z_{i2}y \;=\; d_iN \qquad \text{for } i = 1,\ldots,n. \tag{1}$$

Our first claim is that $\gcd(x, N) = 1$ and $\gcd(y, N) = 1$. Suppose for the sake of contradiction that $\gcd(x, N) = t > 1$. Then $\gcd(t, y) = 1$, as otherwise N is not the least common denominator of s_1 and s_2. Moreover, (1) implies that t divides z_{i2} for all $i = 1,\ldots,n$. But then we could get a better solution by redefining s_2 as $t\,s_2$ and by redefining z_{i2} as z_{i2}/t for $i = 1,\ldots,n$. This contradiction shows that $\gcd(x, N) = 1$. A symmetric argument yields $\gcd(y, N) = 1$.

Fix an arbitrary prime divisor p of N. By the above discussion, p neither divides x nor y. We define q as the unique integer $0 < q < p$ for which p divides $qx - y$. Existence and uniqueness of q follows from elementary number theory. We conclude from (1) that modulo p, we have $0 \equiv z_{i1}x + z_{i2}y \equiv z_{i1}x + z_{i2}qx \equiv x(z_{i1} + qz_{i2})$. Consequently,

$$z_{i1} + qz_{i2} \;\equiv\; 0 \pmod{p} \qquad \text{for } i = 1,\ldots,n. \tag{2}$$

Suppose that for every index i $(1 \le i \le n)$ either $z_{i1} = 0$ or $z_{i2} = 0$ holds. In this case (2) implies that all values z_{ij} are divisible by p. But then multiplying s_1 and s_2 by p and dividing all values z_{ij} by p would yield a better feasible solution. This contradiction demonstrates that

$$\exists\, i: \quad z_{i1} \neq 0 \text{ and } z_{i2} \neq 0. \tag{3}$$

Now we are ready to construct a better feasible solution. We define two new stroke lengths

$$\begin{aligned} s'_1 &= qs_1 - s_2 \\ s'_2 &= (p - q)s_1 + s_2. \end{aligned}$$

Since $s_1 > s_2$ and $0 < q < p$, the numbers s'_1 and s'_2 are again non-negative real numbers. Moreover, for $i = 1,\ldots,n$ we define

$$\begin{aligned} z'_{i1} &= \frac{1}{p}\left(z_{i1} + (q - p)z_{i2}\right) \\ z'_{i2} &= \frac{1}{p}\left(z_{i1} + qz_{i2}\right). \end{aligned}$$

By (2), the numbers z'_{i1} and z'_{i2} are integers. It is easily verified that for every i we have $z_{i1}s_1 + z_{i2}s_2 = z'_{i1}s'_1 + z'_{i2}s'_2$. Therefore,

$$z'_{i1}s'_1 + z'_{i2}s'_2 \;=\; d_i \qquad \text{for } i = 1,\ldots,n.$$

To summarize, the values s'_1, s'_2 together with z'_{ij} for $1 \le i \le n$ and $1 \le j \le 2$ constitute another feasible solution of the golf instance. Now let us compare the objective functions of the original and the new feasible solution. Either $|z'_{i1}| + |z'_{i2}| = |z'_{i1} + z'_{i2}|$, or $|z'_{i1}| + |z'_{i2}| = |z'_{i1} - z'_{i2}|$. We discuss these two cases separately. In the first case

$$|z'_{i1} + z'_{i2}| \;=\; \frac{1}{p}\,|2z_{i1} + (2q - p)z_{i2}| \;\le\; \frac{2}{p}\,|z_{i1}| + \frac{|2q - p|}{p}\,|z_{i2}| \;\le\; |z_{i1}| + |z_{i2}|. \tag{4}$$

Here the first inequality follows from the triangle inequality, and the second inequality follows since $p \ge 2$ and since $0 < q < p$. Moreover, unless $p = 2$ and $z_{i2} = 0$ this second inequality is strict. In the second case

$$|z'_{i1} - z'_{i2}| \;=\; \frac{1}{p}\,|pz_{i2}| \;=\; |z_{i2}| \;\le\; |z_{i1}| + |z_{i2}|. \tag{5}$$

Unless $z_{i1} = 0$, this inequality is strict. In either case, we have shown that $|z'_{i1}|+|z'_{i2}| \leq |z_{i1}|+|z_{i2}|$. By adding up these inequalities for $i = 1, \ldots, n$ we get

$$\sum_{i=1}^{n} \sum_{j=1}^{2} |z'_{ij}| \;\leq\; \sum_{i=1}^{n} \sum_{j=1}^{2} |z_{ij}|. \tag{6}$$

By (3), there exists some index i for which $z_{i1} \neq 0$ and $z_{i2} \neq 0$. By (4) and (5), for this index we have strict inequality $|z'_{i1}| + |z'_{i2}| < |z_{i1}| + |z_{i2}|$. This yields that in (6) in fact strict inequality holds. To summarize, if the least common denominator N of the stroke lengths s_1 and s_2 is divisible by some prime $p \geq 2$, then we can construct another feasible solution s'_1, s'_2, and z'_{ij} for $1 \leq i \leq n$ and $1 \leq j \leq 2$ with a strictly better objective value.

Theorem 3.1. *In the golf problem with $k = 2$ stroke lengths, every optimal solution is integral.*

Consider some instance of the golf problem with $k = 2$ stroke lengths for which every distance d_i is divisible by t. Suppose that there exists some optimal solution with integral stroke lengths s_1 and s_2 that are not divisible by t. Then s_1/t and s_2/t are optimal stroke lengths for the scaled instance with integral distances d_i/t. That is a contradiction, and therefore in all optimal solutions for $k = 2$ the stroke lengths must be divisible by t.

4. The case with three or more stroke lengths

In this section we discuss the case of three and more stroke lengths. The following lemma was useful for cutting down the search space in one of our computer programs.

Lemma 4.1. *For any instance of the golf problem with $k = 3$ stroke lengths, there exist an optimal solution and an optimal integral solution in which $s_j \leq d_{\max}$ holds for all $j = 1, \ldots, k$.*

Proof. Consider an optimal solution (respectively, optimal integral solution) with $s_1 \geq s_2 \geq s_3 > 0$ that has the smallest possible value s_1. Suppose for the sake of contradiction that $s_1 > d_{\max}$. To simplify the presentation of the argument, we flip the signs of z_{i1}, z_{i2}, z_{i3}, d_i for every index i with $z_{i1} < 0$. As a consequence, $z_{i1} \geq 0$ holds for all $i = 1, \ldots, n$. We partition the indices i ($1 \leq i \leq n$) into six classes $C_1, \ldots, C_6$.

C_1: $z_{i1} = 0$; z_{i2} and z_{i3} arbitrary
C_2: $z_{i1} > 0$; $z_{i2} = 0$; $z_{i3} < 0$
C_3: $z_{i1} > 0$; $z_{i2} < 0$; $z_{i3} = 0$
C_4: $z_{i1} > 0$; $z_{i2} > 0$; $z_{i3} < 0$
C_5: $z_{i1} > 0$; $z_{i2} < 0$; $z_{i3} > 0$
C_6: $z_{i1} > 0$; $z_{i2} < 0$; $z_{i3} < 0$

It is easy to see that these six classes cover all possibilities, as $s_1 > d_{\max}$. Moreover, for $i \in C_2 \cup C_4$ we have $|z_{i3}| \geq z_{i1}$, and for $i \in C_3 \cup C_5$ we have $|z_{i2}| \geq z_{i1}$, and for $i \in C_6$ we have $|z_{i2}| + |z_{i3}| \geq z_{i1}$.

We construct a new solution with stroke lengths $s'_1 = s_1 - s_2$, $s'_2 = s_2$, and $s'_3 = s_3$, and with $z'_{i1} = z_{i1}$, $z'_{i2} = z_{i2} + z_{i1}$, and $z'_{i3} = z_{i3}$ for $i = 1, \ldots, n$. It is easy to check that for all $i = 1, \ldots, n$ we have

$$z_{i1}s_1 + z_{i2}s_2 + z_{i3}s_3 \;=\; z'_{i1}s'_1 + z'_{i2}s'_2 + z'_{i3}s'_3.$$

Therefore the new solution is again a feasible solution. Now let us analyze the change Δ_1 in the objective value:

$$\begin{aligned}\Delta_1 &= \sum_{i=1}^{n}(|z'_{i1}|+|z'_{i2}|+|z'_{i3}|-|z_{i1}|-|z_{i2}|-|z_{i3}|) \;=\; \sum_{i=1}^{n}(|z'_{i2}|-|z_{i2}|)\\ &\le \sum_{i\in C_2\cup C_4} z_{i1} - \sum_{i\in C_3\cup C_5} z_{i1} + \sum_{i\in C_6}(|z_{i2}+z_{i1}|-|z_{i2}|).\end{aligned}$$

Here we used $|z_{i2}| \ge z_{i1}$ to bound the terms for $i \in C_3 \cup C_5$. In a completely symmetric way, we can construct a new feasible solution with stroke lengths $s'_1 = s_1 - s_3$, $s'_2 = s_2$, and $s'_3 = s_3$. The corresponding change in the objective value then is

$$\Delta_2 \;\le\; \sum_{i\in C_3\cup C_5} z_{i1} - \sum_{i\in C_2\cup C_4} z_{i1} + \sum_{i\in C_6}(|z_{i3}+z_{i1}|-|z_{i3}|).$$

We will show below that $\Delta_1 + \Delta_2 \le 0$, and this implies that one of Δ_1 and Δ_2 is non-negative. Then the corresponding new feasible solution is again optimal, but has a strictly smaller value s_1. And that contradiction then completes the proof of this lemma.

Hence, it remains to be shown that $\Delta_1 + \Delta_2 \le 0$. We suppose for the sake of contradiction that $\Delta_1 + \Delta_2 > 0$. From the above inequalities we derive that

$$0 \;<\; \Delta_1+\Delta_2 \;\le\; \sum_{i\in C_6}(|z_{i2}+z_{i1}|-|z_{i2}|+|z_{i3}+z_{i1}|-|z_{i3}|).$$

Then at least one term in the sum in the right hand side must be positive, and we have for some $i \in C_6$ that $|z_{i2}+z_{i1}|+|z_{i3}+z_{i1}|-|z_{i2}|-|z_{i3}| > 0$. Since $i \in C_6$, we have $z_{i1} > 0$, $z_{i2} < 0$, and $z_{i3} < 0$. Moreover, we have observed before that $|z_{i2}|+|z_{i3}| \ge z_{i1}$. But these five conditions cannot be fulfilled simultaneously.

∎

Interestingly, the statement in Lemma 4.1 does not carry over to the case of $k = 4$ (or more) stroke lengths: Consider the instance with $d_i = 4$ for $i = 1, \ldots, 100$; $d_i = 5$ for $i = 101, \ldots, 200$; $d_i = 7$ for $i = 201, \ldots, 300$; $d_{301} = 93$; $d_{302} = 95$; $d_{303} = 96$. Setting $s_1 = 4$, $s_2 = 5$, $s_3 = 7$, $s_4 = 100$ yields an objective value of 306. It can be checked that every feasible solution with $s_1, s_2, s_3, s_4 \le 96 = d_{\max}$ has an objective value of at least 307.

Next, we discuss the (non-)integrality of optimal solutions for $k \ge 3$. Consider the instance I_3 with $k = 3$ strokes and the $n = 12$ distances

$$2,\ 3,\ 5,\ 11,\ 11,\ 14,\ 16,\ 17,\ 17,\ 19,\ 21,\ 21.$$

In the fractional solution with $s_1 = 10.5$, $s_2 = 8.5$, $s_3 = 5.5$ every distance can be done with only two strokes. The corresponding objective value equals 24. A computer search based on Lemma 4.1 reveals that the optimal integral solution uses at least 25 strokes.

For $k \ge 4$, let the instance I_k consist of the 12 distances in instance I_3, together with 13 times the distance 10^{6i} for $i = 1, \ldots, k-3$. Hence, instance I_k contains $13k - 27$ distances. We claim that the best integral solution for I_k with k stroke lengths uses at least $13k - 14$ strokes. In the first case we assume that one of the values 10^{6i} with $i = 1, \ldots, k-3$ does not occur as a stroke length. Then the corresponding 13 distances 10^{6i} need at least 26 strokes, and the remaining

$13(k-4)+12$ distances each need at least one stroke. Altogether, this would yield at least $13k-14$ strokes, exactly as we claimed. In the second case we assume that all the values 10^{6i} with $i=1,\ldots,k-3$ do occur as a stroke length. Then the $13(k-3)$ distances that are powers of 10 each need a single stroke. The remaining 12 distances from instance I_3 have to be done with three stroke lengths (the stroke lengths of the form 10^{6i} are useless for them). From instance I_3 we know that they need at least 25 strokes. This case again yields at least $13k-14$ strokes, exactly as we claimed.

To summarize, the best integral solution for the instance I_k uses at least $13k-14$ strokes. On the other hand, it is easily verified that there exists a fractional solution for I_k with at most $13k-15$ strokes.

Theorem 4.2. *For every $k \geq 3$, there exists an instance I_k of the golf problem with k stroke lengths for which every optimal solution is non-integral.*

5. The NP-hardness proof

In this section we show that the golf problem is NP-hard. The NP-hardness proof is done by a reduction from the so-called even-odd partition problem EOP (see Garey and Johnson [1]): Given $2m$ ($m \geq 3$) pairwise distinct positive integers q_i ($i=1,\ldots,2m$) such that $\sum_{i=1}^{2m} q_i = 2Q$ and such that $q_i \leq Q$ for $i=1,\ldots,2m$. Does there exist an index set $J \subset \{1,\ldots,2m\}$ such that $\sum_{i\in J} q_i = Q$ and such that $|J \cap \{2i-1,2i\}| = 1$ for all $i=1,\ldots,m$?

Consider some fixed instance of EOP. Let z be the smallest integer for which $m^z \geq Q+1$ and $z > m$. We construct the following instance of the golf problem: The number of stroke lengths is $k=2m$. There are $n=2m^2+1$ distances. For $i=1,\ldots,m$ the distance $d_{2i-1} = m^{3z}+m^z q_{2i-1}+m^i$ and the distance $d_{2i} = m^{3z}+m^z q_{2i}+m^i$ both occur exactly m times. Note that the $2m$ distances d_i with $1 \leq i \leq 2m$ are pairwise distinct. Finally, the distance d^* equals $m^{3z+1}+m^zQ+\sum_{i=1}^m m^i$.

Lemma 5.1. *If the EOP instance has answer YES, then the golf instance has a solution with at most $2m^2+m$ strokes.*

Proof. Consider the $2m$ stroke lengths d_i with $i=1,\ldots,2m$. Then every distance d_i can be done with a single stroke, which amounts to $2m^2$ strokes altogether. Next, let the index set J be the solution to the EOP instance. It can be verified that $\sum_{i\in J} d_i = m^{3z+1}+m^zQ+\sum_{i=1}^m m^i$. Hence, the distance d^* can be done with $|J|=m$ strokes. ∎

Lemma 5.2. *If the golf instance has a solution with at most $2m^2+m$ strokes, then the EOP instance has answer YES.*

Proof. Consider a solution to the golf instance that makes at most $2m^2+m$ strokes. We first claim that every distance d_i with $i=1,\ldots,2m$ must occur as a stroke length: Otherwise, the m copies of d_i each needed at least two strokes, the remaining $2m^2+1-m$ distances each needed at least one stroke, and this would yield at least $2m^2+m+1$ strokes. Hence, all $2m$ distances d_i indeed occur as stroke lengths. Since these are pairwise distinct numbers, we have fully determined the $2m$ stroke lengths.

Since the distance d^* can be done with m strokes, there exist $2m$ integers z_i with $i=1,\ldots,2m$ such that $\sum_{i=1}^{2m} z_i d_i = d^*$ and $\sum_{i=1}^{2m} |z_i| \leq m$. Let $K \subseteq \{1,\ldots,2m\}$ denote the set of all indices

i with $z_i \geq 0$. Observe that $d_i < m^{3z} + m^z Q + m^m$. Putting things together we get that

$$m^{3z+1} + m^z Q + \sum_{i=1}^{m} m^i \;=\; \sum_{i=1}^{2m} z_i d_i \;\leq\; \sum_{i\in K} z_i d_i \;<\; \sum_{i\in K} z_i (m^{3z} + m^z Q + m^m).$$

Now suppose for the sake of contradiction that $\sum_{i\in K} z_i \leq m-1$. Plugging this in the above inequality and simplifying then implies

$$m^{3z} \;\leq\; (m-2)m^z Q + (m-2)m^m \;<\; (m-2)m^z(Q+1).$$

By our choice of z, we have $m^z \geq m-2$, and $m^z \geq Q+1$, and (trivially) $m^z \geq m^z$. By multiplying these three inequalities, we get the contradiction $m^{3z} \geq (m-2)m^z(Q+1)$. This shows $\sum_{i\in K} z_i \geq m$, and thus *all* integers z_i with $1 \leq i \leq 2m$ are non-negative, and $\sum_{i=1}^{2m} z_i = m$ holds. We now consider the equation $\sum_{i=1}^{2m} z_i d_i = d^*$ modulo m^{m+1}. This yields

$$\sum_{i=1}^{m} (z_{2i-1} + z_{2i})\, m^i \;\equiv\; \sum_{i=1}^{m} m^i \pmod{m^{m+1}}.$$

By considering this equation modulo m^2, modulo m^3, modulo m^4, ..., modulo m^{m+1}, we can show step by step that $z_1 + z_2 = 1$, that $z_3 + z_4 = 1$, that $z_5 + z_6 = 1$, ..., and finally that $z_{2m-1} + z_{2m} = 1$. These arguments use that $\sum_{i=1}^{2m} z_i = m$ and that all z_i are non-negative.

Finally, consider the index set $J \subseteq \{1, \ldots, 2m\}$ that contains all indices i with $z_i = 0$. Since $z_{2i-1} + z_{2i} = 1$ for all $i = 1, \ldots, m$, we have that $|J \cap \{2i-1, 2i\}| = 1$ for all $i = 1, \ldots, m$. Moreover, $\sum_{i=1}^{2m} z_i d_i = d^*$ implies that $\sum_{i\in J} q_i = Q$. Hence, J constitutes the desired solution to the EOP instance.

∎

Computing the golf instance from the EOP instance can easily be done in polynomial time. With this, our Lemmas 5.1 and 5.2 imply the following theorem.

Theorem 5.3. *The golf problem is an NP-hard problem.*

6. Conclusions

In this paper we have established NP-hardness and an integrality result for the golf problem. The most interesting open question probably is to fully understand the integrality gap of this problem. Does there exist an absolute error bound $\alpha > 1$ such that in any instance I the objective value of the best integral solution is at most a factor α above the objective value of the global optimum?

Several questions on the complexity of the golf problem and the k-golf problem remain open. Our NP-hardness proof yields that the general golf problem is NP-hard in the ordinary sense. Is the golf problem also NP-hard in the strong sense? Or is it solvable in pseudo-polynomial time? For the k-golf problem we have constructed a pseudo-polynomial time algorithm. Does there also exist a polynomial time algorithm for the k-golf problem? Or is the k-golf problem NP-hard in the ordinary sense? A first step towards settling these questions might be to understand the special case with $n = 3$ distances and $k = 2$ stroke lengths.

Finally, let us recapitulate and answer the questions on the mathematical puzzle that were posed at the beginning of the introduction: *Are 26 strokes the best possible solution?* — Yes, 26 strokes indeed are the best possible solution. *All nine distances are integer multiples of 25. Does this imply that in any optimal solution x and y must be integer multiples of 25?* — Yes, in any optimal solution x and y must be integer multiples of 25. This follows from the discussion at the end of Section 3. *How does one show that 26 strokes are best possible?* — Since x and y must be integer multiples of 25, we can scale the whole instance by 25; this brings the distances down to 6, 9, 10, 11, 12, 13, 14, 16, 17. By Lemma 4.1 there exists an optimal solution for the scaled instance with $s_1 \leq 17$ and $s_2 \leq 17$. There just remain a few cases, and it is easy to check all these cases by means of a computer program.

References

[1] M.R. Garey and D.S. Johnson [1979]. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco.

[2] H.W. Lenstra Jr. [1983]. Integer programming with a fixed number of variables. *Mathematics of Operations Research 8*, 538–548.

[3] A. Schrijver [1986]. *Theory of Linear and Integer Programming*. John Wiley.