❙ Elemente der Mathematik

Euclid, Calkin & Wilf – Playing with rationals

Stefanie Hofmann, Gertraud Schuster, and Jörn Steuding

Jörn Steuding received his Ph.D. from the University of Hannover in 1999. He received his habilitation at the University of Frankfurt am Main in 2004. Since 2006 he is professor for number theory at the Department of Mathematics of Würzburg University.

Stefanie Hofmann and Gertraud Schuster are teacher students in mathematics and physics at Würzburg University. Their theses written under the supervision of Jörn Steuding motivated this paper (see `www.mathematik.uni-wuerzburg.de/~steuding/sonstiges.html`)

1 Enumerating the rationals à la Cantor

Since Cantor opened the gates of the paradise,[1] we know that the set $\mathbb{Q}$ of rational numbers is countable whereas the set of real numbers is uncountable. The enumeration of $\mathbb{Q}$ follows

---

[1] "No one shall expel us from the Paradise that Cantor has created" are the words of Hilbert about Cantor's foundation of set theory.

Das kombinatorische Spiel *Euklid* wurde 1969 von Cole & Davie eingeführt. Zwei Spieler ziehen abwechselnd aus einer Position $(a, b) \in \mathbb{N}^2$, wobei der ziehende Spieler ein Vielfaches der kleineren Zahl von der grösseren Zahl zu subtrahieren hat, so dass das Resultat noch in $\mathbb{N}^2$ liegt. Der erste Spieler, der nicht mehr ziehen kann, hat verloren. Bislang kennt man elementare (Cole & Davie), geometrische (Lengyel) und zahlentheoretische (Schwartz) Beschreibungen der Gewinnpositionen für *Euklid* (optimales Spiel beider Spieler vorausgesetzt). Die Autoren des nachfolgenden Beitrags fügen dem eine graphentheoretische Gewinnstrategie hinzu, indem sie *Euklid* in Beziehung zum *Calkin-Wilf* Baum setzen. Dieser bemerkenswerte, von Calkin & Wilf im Jahr 2000 eingeführte Baum generiert durch eine einfache Vorschrift alle positiven rationalen Zahlen in gekürzter Form genau einmal.

easily from the following scheme:

$$
\begin{array}{ccccccc}
\frac{1}{1} \longrightarrow \frac{2}{1} & \frac{3}{1} \longrightarrow \frac{4}{1} & \frac{5}{1} \longrightarrow \frac{6}{1} & \cdots \\
\frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \frac{4}{2} & \frac{5}{2} & \cdots \\
\frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{4}{3} & \cdots \\
\frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \cdots \\
\frac{1}{5} & \cdots & \cdots \\
\cdots
\end{array}
$$
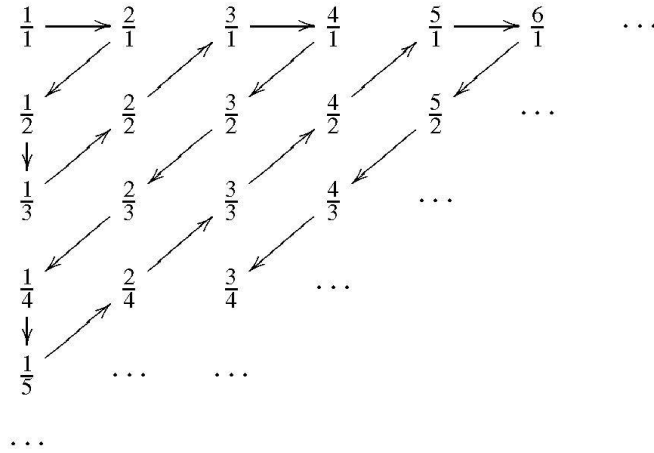
Actually, this scheme gives an enumeration of the set $\mathbb{Q}^+$ of all positive rationals. However, there is a flaw: each element appears infinitely often but only once as a reduced fraction (i.e., with coprime numerator and denominator); the first appearance is in reduced form, all others are integer multiples, forming a straight line.
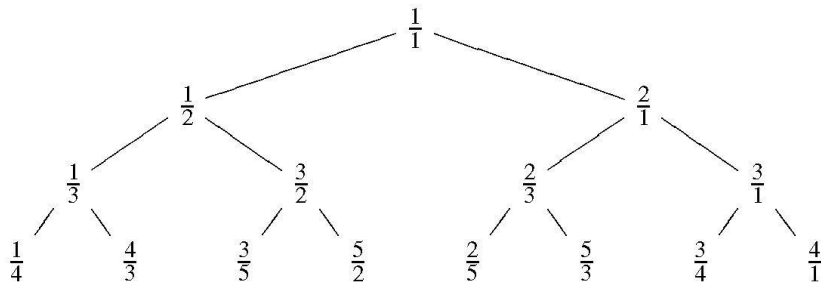
Of course, there are many other enumerations of $\mathbb{Q}^+$ (or $\mathbb{Q}$) known; for a nice overview we refer to Bradley [3]. In this article, we shall investigate the marvellous enumeration of Calkin & Wilf [4] with respect to an application in game theory.

2   The Calkin-Wilf tree

Consider the binary tree generated by the iteration

$$
\frac{a}{b} \;\mapsto\; \frac{a}{a+b}, \quad \frac{a+b}{b}, \tag{1}
$$

starting from the root $\frac{1}{1}$. We call the numbers $\frac{a}{a+b}$ and $\frac{a+b}{b}$ the left and the right child of $\frac{a}{b}$, respectively; we also say that $\frac{a}{b}$ is the mother of its children. We define the notion of generation by induction: the root $\frac{1}{1}$ forms the first generation; the $n+1$-st generation is the set of all children of elements of the n-th generation. The binary tree obtained from this iteration will be called the Calkin-Wilf tree. Here are the first generations:

$$
\begin{array}{c}
\frac{1}{1} \\
\frac{1}{2} \qquad\qquad \frac{2}{1} \\
\frac{1}{3} \quad \frac{3}{2} \qquad \frac{2}{3} \quad \frac{3}{1} \\
\frac{1}{4} \; \frac{4}{3} \quad \frac{3}{5} \; \frac{5}{2} \quad \frac{2}{5} \; \frac{5}{3} \quad \frac{3}{4} \; \frac{4}{1}
\end{array}
$$

Calkin & Wilf [4] have shown that *this tree contains each positive rational number once and only once, each of which represented as a reduced fraction*. This is much better than

the enumeration of $\mathbb{Q}^+$ by Cantor's original procedure; it is also superior to many other enumerations of the rationals (see Bradley [3]). We shall give a sketch of the argument. First of all, each fraction appears in reduced form. To see this, we note that for coprime integers a and b also the linear combination $a + b$ is coprime with a and b, respectively. Since we start from the root $\frac{1}{1}$, coming from the *most* coprime pair of integers, by induction each element of the tree is given in reduced form. For the other claims we define the length of $\frac{a}{b} \in \mathbb{Q}^+$ (with $a, b \in \mathbb{N}$) by $\ell(\frac{a}{b}) = a + b$. Now we want to show that each positive rational eventually appears in the tree. For this purpose we assume that $\frac{a}{b}$ is a positive reduced fraction which does not appear in the tree and which has minimal length among all fractions from $\mathbb{Q}^+$ which do not appear. Then it follows from (1) that also the mother of $\frac{a}{b}$ does not appear. Since the mother has length either b or a, both being less than $\ell(\frac{a}{b}) = a + b$, we get the desired contradiction. Hence, any positive reduced fraction appears at least once. By a similar argument one can show that each positive rational number appears only once (we leave this claim to the interested reader).

Reading the Calkin-Wilf tree line by line, we find

$$\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{1}{3}, \frac{3}{2}, \frac{2}{3}, \frac{3}{1}, \frac{1}{4}, \frac{4}{3}, \frac{3}{5}, \frac{5}{2}, \frac{2}{5}, \frac{5}{3}, \frac{3}{4}, \frac{4}{1}, \cdots \qquad (2)$$

This sequence satisfies the iteration

$$x_1 = 1, \quad x_{n+1} = (2[x_n] + 1 - x_n)^{-1},$$

where $[x]$ denotes the largest integer less than or equal to x. This observation is due to Newman (cf. [9], answering a question of Knuth), resp. Vandervelde & Zagier (cf. [11]).

Actually, the Calkin-Wilf enumeration of $\mathbb{Q}^+$ was already given by Stern fifteen years before Cantor, as recently pointed out by Reznick [10]. In 1858 Stern [12] introduced a sequence $s(n)$ defined by the recursion

$$s(0) = 0, \; s(1) = 1, \; \text{and} \quad s(2n) = s(n), \; s(2n + 1) = s(n) + s(n + 1).$$

He proved that two consecutive elements are coprime and that for any pair a, b of positive coprime integers, there is a unique n such that $a = s(n)$ and $b = s(n + 1)$. It is easy to see that the sequence $\frac{s(n)}{s(n+1)}$ for $n \in \mathbb{N}$ is identical with (2).

The Calkin-Wilf tree has a lot of interesting features. In this note, we are interested in a new application. We will show that this tree serves as a perfect model for a certain Nim-type game.

3 The game Euclid

The game *Euclid* was introduced by Cole & Davie [5]. It is played by two persons, and a position consists of a pair (a, b) of positive integers. The players alternate moves, where a move consists of decreasing the larger number in the current position by any positive multiple of the smaller number, as long as the result is positive. The first player who cannot make a move loses. Cole & Davie have determined those pairs (a, b) which guarantee the first player a win with optimal play.

Theorem 1. *The first player has a winning strategy if and only if the ratio of the larger number to the smaller in the starting position is greater than* $G := \frac{1}{2}(\sqrt{5} + 1)$.

The number G is the golden ratio. There is a nice geometric argument to prove Theorem 1 due to Lengyel [7] which we briefly reproduce here. Define $g := \frac{1}{G} = \frac{1}{2}(\sqrt{5} - 1)$. We consider the open cone

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 : x, y > 0, g < \tfrac{y}{x} < G\}. \tag{3}$$

The goal of *Euclid* is to move to the diagonal $y = x$ since then the other player cannot move. Indeed, for any position (a, b) off the diagonal there is only one direction in the plane in which one can make a move; this is horizontal, if $a > b$, or vertical, otherwise. Further, for every a, there are exactly a points $(x, y) \in \mathcal{C}$ with $x = a$ (this follows from the fact that both g and G are irrational). Thus, if $a < b$ and $(a, b) \notin \mathcal{C}$, then there is a unique positive integer k such that $(a, b - ak) \in \mathcal{C}$. Moreover, given a position (a, b) inside $\mathcal{C}$ with $a \neq b$, there is only one possible legal move which leads outside $\mathcal{C}$; if $a = b$, the player cannot move and has lost. Hence, if the starting position (a, b) lies outside the cone $\mathcal{C}$, then, by optimal play, the first player can always realize a move into $\mathcal{C}$ which then forces the second player to move out from $\mathcal{C}$. The game terminates with a position (a, a) for the second player. This yields the assertion of Theorem 1.

It is not difficult to show that the chance for the first player to win from a randomly chosen position in a large square $(0, N]^2$ is approximately equal to $g = 0.61803\ldots$ on average.

In the next section we shall give a graph-theoretical characterization of all winning positions for the second player based on the Calkin-Wilf tree. Actually, Lengyel [7, 8] suggested to use the related Stern-Brocot tree in order to compute the Sprague-Grundy function for Euclid.

4   A winning strategy based on the Calkin-Wilf tree

If a game starts from a position (a, b) where a and b have a greatest common divisor $d > 1$, then it is easily seen that optimal play for (a, b) is equivalent to optimal play for the reduced position $(a/d, b/d)$. In the sequel we therefore may assume that the starting position (a, b) consists of coprime integers a and b.

We introduce some vocabulary. Given any left child $\frac{a}{b}$ in the Calkin-Wilf tree, we denote by $\mathcal{R}(\frac{a}{b})$ the right subtree which contains all right offsprings of $\frac{a}{b}$ and $\frac{a}{b}$ itself:

$$\frac{a}{b}, \; \frac{a+b}{b}, \; \frac{a+2b}{b}, \; \frac{a+3b}{b}, \; \frac{a+4b}{b}, \; \ldots;$$

we say that $\mathcal{R}(\frac{a}{b})$ is the *right branch* of $\frac{a}{b}$. Similarly, for any right child $\frac{a}{b}$, we denote by $\mathcal{L}(\frac{a}{b})$ the sequence of all left children

$$\frac{a}{b}, \; \frac{a}{a+b}, \; \frac{a}{2a+b}, \; \frac{a}{3a+b}, \; \frac{a}{4a+b}, \ldots$$

and call it the *left branch* of $\frac{a}{b}$. For the root $\frac{1}{1}$, the left and the right branch are given by the numbers $\frac{1}{n}$ and $\frac{n}{1}$ for $n \in \mathbb{N}$, respectively. Obviously, any reduced fraction $\frac{a}{b}$ in the tree is

contained in some left and some right branch which both meet in $\frac{a}{b}$. Thus, each fraction is uniquely determined by two intersecting branches.

Without loss of generality we may assume that $a > b$. To see this, note that the sequence of elements in any fixed generation of the Calkin-Wilf tree is symmetric with respect to the middle (this follows easily from the generating law (1) by induction). Of course, the same symmetry holds for the positions (a, b) in the game *Euclid*. Now assume that we are given a position (a, b). We may think of this position as a reduced fraction in the Calkin-Wilf tree. Since we assume $a > b$, it follows that $\frac{a}{b}$ is a right child (this follows immediately from (1)). Obviously, all possible moves from (a, b) go to $(a - kb, b)$ for some $1 \le k \le [a/b]$. Hence, the possible moves are exactly represented by the ancestors of $\frac{a}{b}$ in the right branch that contains $\frac{a}{b}$. Similarly, if $a < b$, then all possible moves from (a, b) go to $(a, -ka + b)$ for $1 \le k \le [b/a]$ and these are exactly the elements of previous generations in the left branch containing $\frac{a}{b}$.

The path from the quotient of the numbers in the starting position to the root in the Calkin-Wilf tree consists of path segments each of which is part of a left or a right branch (more precisely: any non-empty intersection of the path with a left or right branch is called a line segment). We call a path segment non-trivial if it contains at least two edges, resp. three elements. We enumerate the positions in the game as follows: the starting position is the first position and any move increases the number assigned to the current position by one.

We shall prove the following characterization of the winning positions:

Theorem 2. *The first player has a winning strategy if and only if either the path from the quotient of the numbers in the starting position to the root in the Calkin-Wilf consists only of an odd number of trivial path segments or if the first non-trivial path segment appears at an odd position.*

Proof. We assume that (a, b) with coprime a, b is the starting position. Obviously, the second player wins if the game is started from the position $(1, 1)$. Now suppose that the path in question consists only of trivial path segments. Clearly, since there are no choices for the moves of the players, the elements along this path are alternate winning positions for the two players. Hence, the first player has a winning strategy if and only if the number of these path segments is odd. In the remaining case, there is always at least one non-trivial path segment. Obviously, the player facing a single non-trivial path segment has a choice for his move; he can move to the last or the last but one element in this path segment and, as we have already noticed, exactly one of these two positions is a winning position. Now suppose that there are m non-trivial path segments and $m \in \mathbb{N}$. Then the assertion is proved by induction on m. We assume the statement is true for $m - 1$. By the same reasoning as in the case $m = 1$, the first player facing a non-trivial path segment can move to the last or the last but one element of this path segment. By the right choice, he is the player facing the next non-trivial path segment. By the induction hypothesis, this leads to a winning strategy. This proves Theorem 2.   □

By Theorem 1, all fractions in the Calkin-Wilf tree for which the first player has a winning strategy (as described in the Theorem 2) are less than g or greater than G in value.
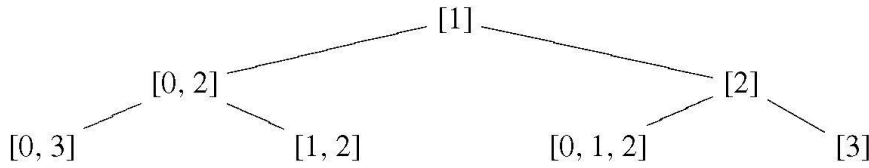
5   Interlude: Continued fractions

It is well-known that each positive rational number x has a representation as a finite (regular) continued fraction

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{\ddots}{\quad + \cfrac{1}{a_{m-1} + \cfrac{1}{a_m}}}}$$

with $a_0 \in \mathbb{N} \cup \{0\}$ and $a_j \in \mathbb{N}$ for $1 \le j \le m$. Actually this is nothing but the *Euclidean algorithm*. In order to have a unique representation, we assume that $a_m \ge 2$ if $m \in \mathbb{N}$. We shall use the standard notation $x = [a_0, a_1, \ldots, a_m]$. In the sequel, we shall also consider some infinite continued fractions, as for example $G = [\overline{1}]$, where $\overline{1}$ indicates the periodic sequence where all partial quotients are equal to 1. Infinite continued fractions represent irrational numbers. For more information about continued fractions and their use in the theory of diophantine approximation we refer to Steuding [13].

Back to the tree! Bird, Gibbons & Lester [2] have shown that *the n-th generation of the Calkin-Wilf tree consists exactly of those rationals having a continued fraction expansion* $[a_0, a_1, \ldots, a_m]$ *for which the sum of the partial quotients a_j is equal to n.* This follows easily from iteration (1) with which the tree was built; notice that this claim is essentially already contained in Lehmer [6] (this was also observed by Reznick [10]).

One can use the approach via continued fractions to locate any positive rational in the tree. Given a reduced fraction x in the Calkin-Wilf tree with continued fraction expansion

$$x = [a_0, a_1, \ldots, a_{m-2}, a_{m-1}, a_m],$$

we set up the path

$$\mathsf{L}^{a_m-1} \mathsf{R}^{a_{m-1}} \mathsf{L}^{a_{m-2}} \cdots \mathsf{L}^{a_1} \mathsf{R}^{a_0} \quad \text{if} \quad m \text{ is odd, and} \tag{4}$$

$$\mathsf{R}^{a_m-1} \mathsf{L}^{a_{m-1}} \mathsf{R}^{a_{m-2}} \cdots \mathsf{R}^{a_1} \mathsf{L}^{a_0} \quad \text{if} \quad m \text{ is even;} \tag{5}$$

note that $a_m - 1 \ge 1$ for $m \in \mathbb{N}$. The notation R^a with $a \in \mathbb{N} \cup \{0\}$ means to go a steps to the right, whereas L^b with $b \in \mathbb{N} \cup \{0\}$ stands for b steps to the left. Then, starting from the root $\frac{1}{1}$ and following this path from left to right, we end up with the element x. It is not difficult to prove this rule. Actually, this law contains much of the mathematics behind the Calkin-Wilf tree.

Back to the game! There is another classification of the winning positions in the game *Euclid* due to R.E. Schwartz (cf. [7]) in terms of the continued fraction expansion. It simply says that the first player who has a choice to move wins:

Theorem 3. *Let (a, b) with $a < b$ be the starting position and denote by $[a_0, a_1, \ldots, a_m]$ the continued fraction expansion of $\frac{b}{a}$. Then the first player has a winning strategy if and only if the first partial quotient a_j that is different from 1 appears at a position with an even index j.*

Hence the winning positions (a, b) with $b > a$ for the second player are exactly represented by those continued fraction expansions $\frac{b}{a} = [a_0, a_1, \ldots, a_m]$, where the first partial quotient $a_j \neq 1$ appears at a position with odd index j. This corresponds via Theorem 2 with the path to $\frac{b}{a}$ which is explicitly given by either (4) or (5).

6   Playing Euclid on the Calkin-Wilf tree

One may have the idea to play Euclid on the Calkin-Wilf tree in the following way: some positive integer n is chosen (randomly) and then a position (a, b) from one of the first n generations of the Calkin-Wilf tree. Alkauskas & Steuding [1] have shown that the mean of the elements of the n-th generation of the Calkin-Wilf tree is tending to $\frac{3}{2}$ as $n \to \infty$. Thus one might think that the second player has better chances in this variant of Euclid since the mean $\frac{3}{2}$ corresponds to a straight line in the cone $\mathcal{C}$, defined by (3), the region of winning positions for the second player. However, this is not true. The mean being equal to $\frac{3}{2}$ does not mean that the elements are clustered around $\frac{3}{2}$. Actually, the converse is true!

We shall give an elementary proof that the first player wins two thirds of the games. The argument does not involve anything about the distribution function as in [1] and, moreover, has the advantage to give an exact expression for the number of wins of the first player provided optimal play.

Denote by f_n and g_n the number of elements in the n-th generation which are a winning position for the first and for the second player, respectively. We observe that

n	1	2	3	4	5	6	7
f_n	0	2	2	6	10	22	42
g_n	1	0	2	2	6	10	22

Clearly,

$$f_n + g_n = 2^{n-1}.$$

It is easy to see that if $\frac{a}{b}$ gives a win for the first player, then also one of its children is a winning position for the first player while the other child yields a win for the second player. If $\frac{a}{b}$ is a winning position for Player 2, both children are a winning position for the first player. To see this we note

$$\frac{a}{b} > G \quad \Longleftrightarrow \quad \frac{a+b}{b} = 1 + \frac{a}{b} > G + 1, \quad g < \frac{a}{a+b} < G,$$

$$g < \frac{a}{b} < G \quad \Longleftrightarrow \quad \frac{a+b}{b} = 1 + \frac{a}{b} > g + 1 = G, \quad \frac{a}{a+b} < g$$

(we leave the easy verification to the interested reader). This leads to the recursion formulae

$$f_{n+1} = f_n + 2g_n \qquad \text{and} \qquad g_{n+1} = f_n \qquad \text{for} \quad n \in \mathbb{N}.$$

Substituting the second into the first, we get

$$f_{n+1} = f_n + 2f_{n-1} \qquad \text{and} \qquad 2^n = f_{n+1} + f_n.$$

Hence, $f_{n+1} = 2^n - f_n = 2^n - (2^{n-1} - f_{n-1})$. By induction

$$f_{n+1} = \sum_{j=0}^{n-1} (-1)^j 2^{n-j}.$$

Evaluating this geometric series, we obtain explicit expressions for f_n and g_n:

Theorem 4. *For $n \in \mathbb{N}$,*

$$f_{n+1} = \frac{2}{3}(2^n - (-1)^n) \qquad and \qquad g_{n+1} = \frac{1}{3}(2^n + (-1)^n 2).$$

Clearly, this shows that the winning chances for the first player are approximately two thirds and for the second player approximately one third of the games.

7   A variation of Euclid

It might be interesting to play the following variant: a pair of positive integers (a, b) is chosen by taking an element $\frac{a}{b}$ from some fixed generation of the Calkin-Wilf tree. Each pair (a, b) corresponds to a mother of two children. As we have already pointed out in the previous section, either both are winning positions for the first player or one is a winning position for the first player and one for the second. The second player may choose which of the children of $\frac{a}{b}$ is the starting position for the first player. To analyze the chances in this game, let us count by F_n the number of mothers for which both children of the n-th generation are winning positions for the first player and denote by G_n the number of mothers with, say, *mixed* children. Then $G_n = 2^{n-2} - F_n$. Obviously, $G_n = g_n = f_{n-1}$, where the numbers g_n and f_n were defined in the previous section. Using $2^{n-2} = f_{n-1} + g_{n-1}$, it follows that $F_n = g_{n-1} = f_{n-2}$. By Theorem 4 it follows that now the winning chances for the second player are two thirds.

References

[1] Alkauskas, G.; Steuding, J.: Statistical properties of the Calkin-Wilf tree and applications to finite continued fractions. Preprint.

[2] Bird, R.; Gibbons, J.; Lester, D.: Functional pearl: enumerating the rationals. *J. Funct. Programming* 16 (2006), 281–291.

[3] Bradley, D.M.: Counting the Positive Rationals: A Brief Survey. *Preprint* available at http://germain.umemat.maine.edu/faculty/bradley/papers/pub.html

[4] Calkin, N.; Wilf, H.: Recounting the rationals. *Amer. Math. Monthly* 107 (2000), 360–363.

[5] Cole, A.J.; Davie, A.J.T.: A game based on the Euclidean algorithm and a winning strategy for it. *Math. Gaz.* 53 (1969), 354–357.

[6] Lehmer, D.H.: On Stern's diatomic series. *Amer. Math. Monthly* 36 (1929), 59–67.

[7] Lengyel, T.: A nim-type game and continued fractions. *Fibonacci Quart.* 41 (2003), 310–320.

[8] Lengyel, T.: *On calculating the Sprague-Grundy function for the game Euclid.* Proceedings of the 11th International Conference on Fibonacci Numbers and Their Applications 2004, 169–175 (to appear).

[9] Newman, M.: Recounting the rationals, Continued. Credited in *Amer. Math. Monthly* 110 (2003), 642–643.

[10] Reznick, B.: Regularity properties of the Stern enumeration of the rationals. *Preprint* available at http://de.arxiv.org/pdf/math.NT/0610601.pdf?front

[11] Sloane, N.; Plouffe, S.: *The On-line Encyclopedia of integer sequences.* http://www.research.att.com/~njas/sequences/

[12] Stern, M.A.: Über eine zahlentheoretische Funktion. *J. Reine Angew. Math.* 55 (1858), 193–220.

[13] Steuding, J.: *Diophantine Analysis.* CRC Press/Chapman-Hall 2005.

Stefanie Hofmann
Tulpenstraße 12
D–97950 Gerchsheim, Germany

Gertraud Schuster
Adamweg 5a
D–84036 Landshut, Germany

Jörn Steuding
Institut für Mathematik
Universität Würzburg
Am Hubland
D–97218 Würzburg, Germany
e-mail: steuding@mathematik.uni-wuerzburg.de