Euclidean Strings

Frank Ruskey, U. Victoria, Victoria, CANADA
(visiting U. Newcastle, Newcastle, AUSTRALIA)
John Ellis, U. Victoria, Victoria, CANADA
Joe Sawada, U. Victoria, Victoria, CANADA
Jamie Simpson, Curtin U., Perth, AUSTRALIA

For every rational number there is a string with very nice properties.

Consider strings of natural numbers. Let $p = p_1 p_2 \cdots p_n$, where $k = \sum p_i$.

Main Result: The string p is a *Euclidean string* if and only if one of the following conditions holds.

- The string p is a rotation of $(p_1{+}1)p_2 \cdots p_{n-1}(p_n{-}1)$. E.g.

$$01011 = \text{ShiftLeft}^2(\ 11010\ ).$$

- The string p is lexicographically maximum among all Lyndon words of length n whose characters sum to k, with $\gcd(n,k) = 1$. E.g.,

$$01011 = \max\{00003, 00012, 00021, 00102, 00201, 00111, 01011\}$$

- The string p is the reversal of the histogram of a rational Beatty sequence. E.g.,

$$\mathit{index}(\ 01011\ ) = \{2,4,5\} = 5 - \{0,1,3\}, \quad \text{where}$$

$$\{0,1,3\} = \{\lfloor 0/3 \rfloor, \lfloor 5/3 \rfloor, \lfloor 10/3 \rfloor\} = \{\lfloor 5j/3 \rfloor : j = 0,1,2\}$$

- The string is created from 0 by the rules $x \to x+1$ and $x \to 01^x$. E.g.,

$$0 \to 1 \to 01 \to 12 \to 01011$$

- The string is created from $E_{1,0} = 0$ and $E_{1,1} = 1$ by the rule

$$E_{n,k} = E_{d,c} E_{n-d,k-c},$$

where $dk = cn + 1$ (produces binary strings only). E.g.,

$$E_{5,3} = E_{2,1}E_{3,2} = (E_{1,0}E_{1,1})(E_{2,1}E_{1,1}) = 01(E_{2,1}1) = 01(011) = 01011$$

Let $\mathbf{p} = p_1p_2\cdots p_n$. Define

$$\tau(\mathbf{p}) := (p_1{+}1)p_2\cdots p_{n-1}(p_n{-}1)$$

$$\rho(\mathbf{p}) := p_np_1\cdots p_{n-1}$$

A string is Euclidean if there is an integer d such that

$$\rho^d(\mathbf{p}) = \tau(\mathbf{p})$$

E.g., if $\mathbf{p} = 01011$, then $\rho^2(\mathbf{p}) = 11010 = \tau(\mathbf{p})$.
In general,

$$p_j = \begin{cases} p_1 + 1 & \text{if } j = n - d + 1 \\ p_n - 1 & \text{if } j = n - d \\ p_{j+d} & \text{otherwise} \end{cases}$$

Consider

$$S = p_1, p_{1+d}, p_{1+2d}, \ldots, p_{1+(r-1)d}, p_{1+rd} = p_1.$$

Observe that p_n must be in S:

$$S = p_1, p_{1+d}, p_{1+2d} \ldots, p_{n-d}, p_n, \ldots, p_{n-d+1}, p_1.$$

Thus $1{+}jd = n$ for some j, and thus (with some thought) $\gcd(n, d) = 1$.
Thus S contains all the elements of $\mathbf{p}$.
There are only two values in S, differing by one.

$$S = \underbrace{p_1, p_{1+d}, \ldots, p_{n-d}}_{\lfloor k/n \rfloor}, \underbrace{p_n, \ldots, p_{n-d+1}}_{\lceil k/n \rceil}, p_1.$$

The Symmetric Euclidean Algorithm

procedure gcd(n, k : $\mathbb{N}$) **returns** $\mathbb{N}$;
    $k < n$: return(gcd($n - k$, k));
    $k = n$: return(n);
    $k > n$: return(gcd(n, $k - n$));

procedure Egcd(n, k : $\mathbb{N}$) **returns** String;
    $k < n$: return(exp(Egcd($n - k, k$));   { $exp(x) = 01^x$ }
    $k = n$: return('1');
    $k > n$: return(inc(Egcd($n, k - n$));   { $inc(x) = x + 1$ }

Example:
Egcd(5,3) = exp(Egcd(2,3)) = exp(inc(Egcd(2,1))) = exp(inc(exp(Egcd(1,1))))
= exp(inc(exp(1))) = exp(inc(01)) = exp(12) = 01011.

The Stern-Brocot Tree

L = nearest left ancestor = k_L/n_L
R = nearest right ancestor = k_R/n_R

$$\text{At the node: } \frac{k_L + k_R}{n_L + n_R}$$

To get strings concatenate: LR
Initially 0/1 and 1/1 (numbers) or 0 and 1 (strings)
From this is follows that $E_{n,k} = E_{d,c}E_{n-d,k-c}$, where $kd = nc + 1$.

Lyndon words

A *Lyndon word* is a string that is (strictly) lexicographically smaller than all of its proper rotations.

Note that

$$\rho^i(\mathbf{p}) = p_{n-i+1} \cdots p_{n-1} \underbrace{p_n p_1} p_2 \cdots p_{n-i}$$

$$\begin{aligned} \rho^{i+d}(\mathbf{p}) &= \rho^i(\rho^d(\mathbf{p})) \\ &= \rho^i(\tau(\mathbf{p})) \\ &= \rho^i(p_n p_2 \cdots p_{n-1} p_1) \\ &= p_{n-i+1} \cdots p_{n-1} \underbrace{p_1 p_n} p_2 \cdots p_{n-i-1} \end{aligned}$$

Thus

$$\mathbf{p} \prec \rho^d(\mathbf{p}) \prec \rho^{2d}(\mathbf{p}) \prec \cdots \prec \rho^{(n-1)d}(\mathbf{p})$$

Example:

$$01011 \prec 01101 \prec 10101 \prec 10110 \prec 11010$$

Another related result:

Euclidean strings are lexicographically maximum among all Lyndon words of the same length and weight.