Optimale Basissysteme

von

Ulrike Brandt (Darmstadt)
Gerd Hofmeister (Mainz)

TI-4/89
Institut für Theoretische Informatik

1. Vorbemerkung:

Es sei $A = \{a_1, a_2, \ldots\} \subseteq \mathbf{N}$ mit $1 = a_1 < a_2 < \ldots$ ein Basissystem und $[1, x]$ ein Zahlbereich ($x \geq 1$). Jede ganze Zahl $n \in [1, x]$ besitzt (mindestens) eine Darstellung

$$n = \sum_{a_i \in A} n_i a_i, \quad n_i \in \mathbf{N}_0,$$

bei der $t(n, A) := \sum_{a_i \in A} n_i$ minimal ist; eine solche Darstellung heißt auch Minimaldarstellung von n bezüglich A. Wir setzen

$$T(x, A) := \max\{t(n, A)/n \leq x\}$$

und wie üblich

$$A(x) := \sum_{1 \leq a_i \leq x} 1.$$

Beim Löschen des Inhalts von Registern eines Rechners (Reset-Operation) geht es zum einen um das Auffinden der aktivierten Zellen und zum anderen um die Desaktivierung dieser Zellen. Bei einem zugrundegelegten Basissystem A und einem darzustellenden Zahlbereich $[1, x]$ ist der Aufwand für den ersten Vorgang proportional der benötigten Registerlänge und damit proportional der Anzahl $A(x)$. Den Aufwand für den zweiten Prozeß setzen wir proportional zur Größe $T(x, A)$ an.

Die folgenden Beispiele zeigen, daß zwischen den Maßen $A(x)$ und $T(x, A)$ ein Trade-Off besteht:
Beispiel 1. $A = \{1\}$. Dann ist $A(x) = 1$ und $T(x, A) = x$.
Beispiel 2. $A = \mathbf{N}$. Hier ist $A(x) = x$ und $T(x, A) = 1$.
Ein naheliegendes Maß zur Beurteilung von Registerarchitekturen bei der Behandlung dieses Trade-Off ist das Produkt $A(x) \cdot T(x, A)$. Demgemäß untersuchen wir im folgenden die Frage nach dem Verhalten von

$$m(x) := \min_{A \in P_1} A(x) \cdot T(x, A),$$

wo P_1 die Menge aller $A \subseteq \mathbf{N}$ mit $1 \in A$ bezeichnet. Von besonderem Interesse ist dabei die Frage, wie Mengen B aussehen, für die

$$B(x) \cdot T(x, B) = m(x)$$

gilt. Derartige Mengen sollen $m(x)$-optimal heißen.

In dieser allgemeinen Form scheint das Problem zur Zeit kaum lösbar zu sein; es hängt eng mit dem ebenfalls noch weitgehend ungelösten Reichweitenproblem zusammen, welches auf Rohrbach [8] zurückgeht. Wir können in dieser Arbeit nur Abschätzungen für $m(x)$ angeben, nämlich

$$0.52 \log^2 x \leq m(x) \leq 1.08 \log^2 x \text{ für } x \geq x_0.$$

Anders sieht es aus, wenn man sich anstelle von P_1 auf Teilbereiche $Q_1 \subseteq P_1$ beschränkt und

$$m(x, Q_1) := \min_{A \in Q_1} A(x) \cdot T(x, A)$$

betrachtet. Für $g \geq 2$ setzen wir

$$B_g := \{g^i / i \in \mathbf{N}_0\}$$

und

$$Q_1 := \{B_g / g \in \mathbf{N}, g \geq 2\}.$$

In diesem Fall läßt sich $m(x, Q_1)$ für hinreichend große x bestimmen und B_5 ist dabei $m(x, Q_1)$-optimal. Auch in einem deutlich größeren Bereich, den später zu definierenden regulären Mengen, läßt sich das Optimierungsproblem lösen.

Am Schluß der Arbeit untersuchen wir noch eine Art Gegenfrage, nämlich

$$M(x) := \max_{A \in P_1} A(x) \cdot T(x, A)$$

und zeigen u.a.

$$M(x) = \lfloor \frac{x^2 + 2x + 1}{4} \rfloor \text{ für alle } x \in \mathbf{N}.$$

2. Zum Beweis einer unteren Abschätzung für $m(x)$ zeigen wir

Lemma 1:

Für beliebige $h, k \in \mathbf{N}$ gilt

$$\binom{h+k}{k} \leq \sqrt{\frac{h+k}{2\pi hk}} \cdot 4^{\sqrt{hk}}$$

Beweis: Wir benutzen ein Resultat von Robbins [7] zur Stirlingschen Formel, nämlich

$$\sqrt{2\pi n}\,(\frac{n}{e})^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\,(\frac{n}{e})^n e^{\frac{1}{12n}}$$

für alle $n \in \mathbf{N}$. Damit folgt sofort

$$\binom{h+k}{k} \leq \sqrt{\frac{h+k}{2\pi hk}}\,\frac{(h+k)^{h+k}}{h^h k^k} = \sqrt{\frac{h+k}{2\pi hk}}\,(1+\frac{k}{h})^h (1+\frac{h}{k})^k.$$

Wir setzen $t := \frac{h}{k}$ und können wegen der Symmetrie in h und k o.B.d.A. $h \geq k$, also $t \geq 1$ voraussetzen.

Zum Beweis der Behauptung genügt es zu zeigen

$$f(t) := (1+\frac{1}{t})^t (1+t) \leq 4^{\sqrt{t}} \text{ für } t \geq 1.$$

Für $t \geq 3$ folgt die Ungleichung sofort wegen $(1+\frac{1}{t})^t < e$.

Sei daher $1 \leq t \leq 3$. Wäre $f(t_0) < 4^{\sqrt{t_0}}$ für ein $t_0 \in]1, 3[$, so auch

$$\log f(t_0) > \sqrt{t_0} \log 4.$$

Wir setzen

$$g(t) := \log f(t) - \sqrt{t} \log 4$$

und beachten

$$g(1) = 0, g(t_0) > 0, g(3) < 0.$$

Also existiert ein $t_1 \in]t_0, 3[$ mit $g(t_1) = 0$. Nach dem Satz von Rolle existiert dann ein $t_2 \in]1, t_1[$ mit $g'(t_2) = 0$. Außerdem rechnet man nach, daß $g'(1) = 0$ gilt, also existiert abermals nach dem Satz von Rolle ein $t_3 \in]1, t_2[$ mit $g''(t_3) = 0$.

Es ist aber

$$g''(t) = -\frac{1}{t^2 + t} + \frac{\log 2}{2t\sqrt{t}} = 0$$

gleichwertig mit

$$t^2 - 2(\frac{2}{\log^2 2} - 1)t + 1 = 0,$$

also $t = 6.16\ldots$ oder $t = 0.16\ldots$, mithin $t \notin]1, t_2[$ ($\subseteq]1, 3[$).

Also ist $f(t) \leq 4^{\sqrt{t}}$ für alle $t \geq 1$. ∎

Satz 1: Für alle $x \in \mathbf{N}$ gilt

$$m(x) > \frac{1}{\log^2 4} \log^2 x$$

Beweis: Sei $A = \{1 = a_1 < a_2 < \ldots\} \in P_1$ und $x \in \mathbf{N}$ beliebig.

Wir setzen $k := A(x)$ und $h := T(x, A)$. Dann gilt für die Menge

$$M := \{(n_1, \ldots n_k)/n_i \in \mathbf{N}_0, \sum_{i=1}^{k} n_i \leq h\} : \ |M| = \binom{h+k}{k}.$$

Zu jeder Zahl $y \in [0, x]$ gibt es dann mindestens ein k-tupel $(n_1, \ldots, n_k) \in M$, so daß $\sum_{i=1}^{k} n_i a_i$ eine Darstellung von y ist. Es folgt mit Lemma 1

$$x + 1 \leq \binom{h+k}{k} \leq 4^{\sqrt{hk}},$$

d.h. $A(x)T(x, A) = kh > \dfrac{\log^2 x}{\log^2 4}$. ∎

Wir erinnern daran, daß Q_1 die Menge aller g-adischen Basissysteme für $g = 2, 3, \ldots$ bezeichnet und zeigen

Satz 2:

Es gilt

$$m(x, Q_1) = \frac{4}{\log^2 5} \log^2 x + 0(\log x),$$

also

$$m(x, Q_1) < 1.545 \log^2 x \text{ für } x > x_0,$$

und spätestens für alle $x > 10^{337}$ ist $B_5 = \{1, 5, 5^2, \ldots\}$ $m(x, Q_1)$-optimal.

Beweis: Es seien $g, x \in \mathbf{N}$ mit $g \geq 2$. Dann existiert ein $k \in \mathbf{N}$ mit

$$g^{k-1} \leq x < g^k$$

und man erhält

$$B_g(x) = k = \lfloor \frac{\log x}{\log g} \rfloor + 1.$$

Zur Bestimmung von $T(x, B_g)$ stellt man fest, daß

$$n_0 := \sum_{i=0}^{k-2} (g-1)g^i + yg^{k-1}$$

mit $y := \lfloor \frac{x+1}{g^{k-1}} \rfloor - 1 \leq g-1$ unter allen Zahlen $n \leq x$ maximal viele Summanden benötigt, d.h.

$$(g-1)(k-1)k \leq B_g(x)T(x, B_g) \leq (g-1)k^2$$

oder

$$\begin{aligned} f_1(x,g) &:= \frac{g-1}{\log^2 g} \log^2 x - \frac{g-1}{\log g} \log x \\ &\leq B_g(x)T(x, B_g) \\ &\leq \frac{g-1}{\log^2 g} \log^2 x + 2\frac{g-1}{\log g} \log x + g - 1 =: f_2(x,g). \end{aligned}$$

Eine Extremalbetrachtung zeigt, daß $f_2(x,5) < f_1(x,g)$, sofern $x \geq g^2$ und $x > 10^{337}$.

Für $x < g$ ist $B_g(x)T(x, B_g) = x$ und für $g \leq x < g^2$ gilt $B_g(x)T(x, B_g) \geq 2(g-1)$. In beiden Fällen zeigt man, daß $f_2(x,5) < B_g(x)T(x, B_g)$ für $g \neq 5$ und $x > 10^6$ gilt. ∎

In diesem Zusammenhang verweisen wir auch auf die Arbeit von Hofmeister und Waadeland [5].

Definition:

Es sei $A = \{1 = a_1 < a_2 < \ldots\} \subseteq \mathbf{N}$ eine endliche oder unendliche Zahlenmenge und $x \in \mathbf{N}$. Eine Darstellung

$$x = \sum_{i=1}^{k} r_i a_i, \; r_i \in \mathbf{N}_0$$

heißt regulär, wenn

(i) $a_k \leq x < a_{k+1}$ (falls ein solches Element $a_{k+1} \in A$ überhaupt existiert)

(ii) $\sum_{i=1}^{j} r_i a_i < a_{j+1}$ für alle $j = 1, \ldots, k-1$

(vgl. Hofmeister [3]).

Jedes x besitzt genau eine reguläre Darstellung bez. A. Wie in [3] setzen wir $r(x, A) := \sum_{i=1}^{k} r_i$, falls $\sum_{i=1}^{k} r_i a_i$ die reguläre Darstellung von x bezüglich A ist und $R(x, A) := \max\{r(y, A)/y \leq x\}$. Die reguläre

Darstellung von x ist nicht notwendig eine Minimaldarstellung von x bzgl. A, wie das einfachste Beispiel $A = \{1, 3, 4\}$ und $x = 6$ zeigt.

Definition:

Ist für jedes $x \in \mathbf{N}$ die reguläre Darstellung von x bzgl. A eine Minimaldarstellung von x, so heißt die Menge A regulär.

Djawadi [2] verwendet stattdessen den Begriff "angenehm" und gibt Kriterien für derartige Mengen.

Es bezeichne R_1 das System der regulären Mengen aus P_1; insbesondere gilt

$$Q_1 \subset R_1 \subset P_1.$$

Hofmeister [3] gibt für $A \in R_1$ einen Algorithmus an zur Bestimmung von

$$g(h, A) := \max\{x \in \mathbf{N} / R(x, A) \leq h\}$$

für beliebige $h \in \mathbf{N}$ und Mrose [6] bestimmt

$$g(h, k) := \max\{g(h, A) / A \in R_1 \text{ und } |A| = k\}$$

einschließlich der zugehörigen $g(h, k)$-optimalen (regulären) Mengen.

Satz 3:

Mit $\lambda = \dfrac{3 + \sqrt{5}}{2}$ gilt

$$m(x, R_1) < 1.08 \log^2 x \text{ für } x > x_0.$$

Beweis: Es gilt

$$\begin{aligned} m(x, R_1) &= \min\{hk / g(h, k) \geq x\} \\ &\leq \min\{hh / g(h, h) \geq x\}. \end{aligned}$$

Nach Hofmeister [4] ist

$$g(h, h) \geq c\lambda^h - 1 \text{ mit } c = \frac{5 + \sqrt{5}}{10}.$$

Für $h = \left\lceil \dfrac{\log(x + 1) - \log c}{\log \lambda} \right\rceil$ folgt daher

$$g(h, h) \geq x,$$

also

$$m(x, R_1) \leq \left\lceil \frac{\log(x + 1) - \log c}{\log \lambda} \right\rceil^2$$

und damit die Behauptung. ∎

Bemerkung:

Mit deutlichem Mehraufwand zeigt Brandt [1] darüberhinaus

$$m(x, R_1) = \frac{\log^2 x}{\log^2 \lambda} + 0(\log x)$$

und bestimmt $m(x, R_1)$-optimale Basissysteme. Für unendlich viele x sind dies sowohl die Fibonacci-Zahlen mit geradem Index als auch die mit ungeradem.

Mit den Sätzen 1 und 3 hat man insbesondere die

Folgerung:

Für $x > x_0$ gilt

$$0.52 \log^2 x \leq m(x) \leq 1.08 \log^2 x.$$

3. Interessant erscheint uns auch die Gegenfrage nach den ineffektivsten Mengen A (mit $1 \in A$), die man in das Intervall $[1, x]$ packen kann. Dazu setzen wir

$$M(x, \Omega) = \max_{A \in \Omega} A(x) T(x, A)$$

und bezeichnen die zugehörigen Mengen A als $M(x, \Omega)$-pessimal. Im Fall $\Omega = P_1$ setzen wir $M(x) = M(x, P_1)$.

Satz 4: Sei $x \in \mathbf{N}$. Dann gilt
$M(x) = \lfloor \frac{x^2 + 2x + 1}{4} \rfloor$ und zugehörige $M(x)$-pessimale Mengen sind

$$B_1 = \{1\} \cup [\frac{x+3}{2}, x] \text{ für ungerades } x, \text{ bzw.}$$
$$B_{2,1} = \{1\} \cup [\frac{x+2}{2}, x] \text{ und}$$
$$B_{2,2} = \{1\} \cup [\frac{x+4}{2}, x] \text{ für gerades } x.$$

Beweis: Sei $A \subseteq [1, x]$. Dann existieren s, $y \in [1, x]$ mit $t(y, A) = T(x, A) = s$.

Sei zunächst $s \geq 2$. Dann folgt für die $s-1$ Elemente $y-(s-2), y-(s-3), \ldots, y \notin A$, weil y andernfalls eine Darstellung $y = 1 \cdot (y - j) + j \cdot 1$ mit $y - j$, $1 \in A$ besäße und $t(y, A) \leq j + 1 \leq s - 1$ wäre.

Setzt man dann

$$B = \begin{cases} \{1\} \cup [s+1, x], & \text{falls } 1 \leq s \leq x - 1 \\ \{1\}, & \text{falls } s = x, \end{cases}$$

so zeigt ein Vergleich mit der Menge A, daß $B(x) \geq A(x)$ und $T(x, B) = T(x, A)$. (Im Fall $s = 1$ oder $s = x$ ist $A = B$.)

Man erhält $A(x)T(x, A) \leq B(x)T(x, B) = (x - s + 1) \cdot s = \frac{(x+1)^2}{4} - (\frac{x+1}{2} - s)^2$ und hiermit folgt leicht die Behauptung. ∎

Satz 5: Für alle $A \in P_1$ und alle $x \in \mathbf{N}$ gilt $A(x)T(x, A) \leq 2\frac{x^2}{A(\sqrt{x})}$

Beweis: Sei $y \leq x$ beliebig und $a \in A$ maximal mit $a \leq \sqrt{x}$. Dann besitzt y bezüglich A die Darstellung

$$y = \lfloor \frac{y}{a} \rfloor a + r \cdot 1,$$

wobei $0 \leq r \leq a - 1$.

Es folgt

$$T(x, A) \leq \frac{x}{a} + a - 1 < 2\frac{x}{a} \leq 2\frac{x}{A(\sqrt{x})},$$

denn $a \geq A(\sqrt{x})$ aufgrund der Maximalitätsbedingung von a.

Wegen $A(x) \leq x$ folgt die Behauptung. ∎

Andererseits existieren unendliche Mengen A und unendliche Folgen $(x_n)_{n \in \mathbf{N}}$ mit

$$A(x_n)T(x_n, A) \geq x_n^{2 - \frac{1}{f(n)}}$$

für jedes $\epsilon > 0$ und $n > n_\epsilon$. Genauer gilt

Satz 6:

Zu jedem $f : \mathbf{N} \to \mathbf{N}$ existiert eine unendliche Menge A und eine gegen ∞ strebende Folge natürlicher Zahlen $(x_n)_{n \in \mathbf{N}}$ mit

$$A(x_n)T(x_n, A) \geq x_n^{2 - \frac{1}{f(n)}}.$$

Beweis: Wir wählen $c_1 = 1$ und $c_n \in \mathbf{N}$ mit $c_n \geq \frac{1}{2}(8c_{n-1})^{f(n)}$ für $n \geq 2$.

Dann folgt

$$8c_{n-1} \leq (2c_n)^{\frac{1}{f(n)}}.$$

Wir wählen weiter

$$A = \{1\} \cup \bigcup_{n=1}^{\infty}]c_n, 2c_n]$$

und $x_n = 2c_n$ für alle $n \in \mathbf{N}$.

Dann folgt

$$A(2c_n) \geq 2c_n - c_n = c_n.$$

Wegen

$$t(c_n, A) \geq \frac{c_n}{2c_{n-1}}$$

folgt weiterhin

$$T(2c_n, A) \geq \frac{c_n}{2c_{n-1}}.$$

Insgesamt erhält man also

$$A(x_n)T(x_n, A) = A(2c_n)T(2c_n, A) \geq \frac{c_n^2}{2c_{n-1}} = \frac{(2c_n)^2}{4(2c_{n-1})} \geq (2c_n)^{2 - \frac{1}{f(n)}} = x_n^{2 - \frac{1}{f(n)}}.$$

∎

Literaturverzeichnis

[1] U. Brandt,
Über den Einfluß der Zahldarstellung auf registerorientierte Rechner.
Habilitationsschrift am Fachbereich Informatik der Technischen Hochschule Darmstadt, 1988.

[2] M. Djawadi,
Kennzeichnung von Mengen mit einer additiven Minimaleigenschaft,
J. reine angew. Math. **311/312** (1979), 307–314.

[3] G. Hofmeister
Über eine Menge von Abschnittsbasen,
J. reine angew. Math. **213** (1963), 43–57.

[4] G. Hofmeister
Über eine Menge von Abschnittsbasen II,
Det Kgl Norske Vidensk. Selsk. Skr. 1966 Nr. 10.

[5] G. Hofmeister, H. Waadeland,
Eine Minimaleigenschaft des Fünfer-Systems,
Det Kgl Norske Vidensk Selsk. Skr. 1966 Nr. 11.

[6] A. Mrose,
Die Bestimmung der extremalen regulären Abschnittsbasen mit Hilfe einer Klasse von Kettenbruchdeterminanten,
Dissertation an der Freien Universität Berlin, 1969.