Zadanie 1. 1 punkt

Udowodnij, że dodatnia liczba rzeczywista ma skończone rozwinięcie dwójkowe wtedy i tylko wtedy, gdy jest postaci $\frac{m}{2^n}$, gdzie m i n są liczbami naturalnymi.

Dowód <=

Załóżmy, że x jest dodatnią liczbą rzeczywistą postaci $\frac{m}{2^n}$, gdzie m i n są liczbami naturalnymi. Chcę dowieść, że x ma skończone rozwinięcie dwójkowe.
Dowód przez indukcję względem n.

Podstawa: $n = 0$
Wtedy $x = \frac{m}{2^0} = \frac{m}{1} = m$, czyli x jest liczbą naturalną.
Stąd $\exists u \in N\colon x = \sum_{i=0}^{u} x_i * 2^i \, , \forall x_i \in \{0,1\}$, czyli x ma skończone rozwinięcie dwójkowe.

Krok: Załóżmy, że $\frac{m}{2^n}$ ma skończone rozwinięcie dwójowe a $x = \frac{m}{2^{n+1}}$.
Wtedy $x = \frac{m}{2^{n+1}} = \frac{m}{2^n} * \frac{1}{2^1}$ również ma SRD, bo iloczyn liczb SRD daje liczbę SRD.
Zatem każda liczba postaci $\frac{m}{2^n}$ ma SRD.

Dowód =>

Załóżmy, że x jest dodatnią liczbą rzeczywistą mającą skończone rozwinięcie dwójkowe. Wiemy, że $x = \lfloor x \rfloor + \{x\}$ gdzie $\lfloor x \rfloor \in N$ oraz $\{x\} \in [0;1)$.
Zatem wystarczy pokazać, że zarówno $\lfloor x \rfloor$ jak i $\{x\}$ można zapisać w postaci $\frac{m}{2^n}$:

$$\lfloor x \rfloor = \frac{\lfloor x \rfloor}{1} = \frac{\lfloor x \rfloor}{2^0} = \frac{m}{2^n} \text{ dla } m = \lfloor x \rfloor \text{ (bo obie } \in N), n = 0$$

Skoro $\{x\} \in [0;1)$ i ma SRD, to $\exists u \in N$ takie, że:

$$\{x\} = \sum_{i=1}^{u} \left(x_i * 2^{-i}\right) \forall x_i \in \{0,1\}$$

Stąd $\{x\} = \sum_{i=1}^{u}\left(x_i * 2^{-i}\right) = \sum_{i=1}^{u}\left(x_i * 2^{-u+u} * 2^{-i}\right) =$

$$= \sum_{i=1}^{u}\left(x_i * 2^{-u} * 2^{u-i}\right) = \frac{\sum_{i=1}^{u}\left(x_i * 2^{u-i}\right)}{2^u} = \frac{m}{2^n} \text{ dla } m = \sum_{i=1}^{u}\left(x_i * 2^{u-i}\right), n = u \blacksquare$$

Zadanie 2. 1 punkt

Ustalmy liczbę $\mathrm{B} \in \{2, 3, 4, \dots\}$. Pokaż, że każda niezerowa liczba rzeczywista x ma jednoznaczne przedstawienie w postaci
$\mathrm{x} = \mathrm{smB^c}$, gdzie $\mathrm{s} = \mathrm{sgn(x)}$, $\mathrm{c} \in \mathrm{Z}$, $\mathrm{m} \in \left[\frac{1}{\mathrm{B}}, 1\right)$.

Na początku zauważmy, że zawsze zachodzi $(*)\ \mathrm{mB^c} > 0, \mathrm{bo\ m} > 0 \mathrm{\ i\ B^c} > 0$
Mamy pokazać, że dla dowolnych $\mathrm{c_1, c_2, m_1, m_2}$ zachodzi

$$\mathrm{x} = \mathrm{sm_1B^{c_1}} = \mathrm{sm_2B^{c_2}} \leftrightarrow \mathrm{c_1} = \mathrm{c_2} \mathrm{\ oraz\ } \mathrm{m_1} = \mathrm{m_2}$$

Dowód nie wprost
Załóżmy, że (Z) $\mathrm{x} = \mathrm{sm_1B^{c_1}} = \mathrm{sm_2B^{c_2}}\ dla\ \mathrm{c_1} \neq \mathrm{c_2} \mathrm{\ oraz\ } \mathrm{m_1} \neq \mathrm{m_2}$
Możemy podzielić (Z) obustronnie przez s dzięki (*):

$$\mathrm{m_1B^{c_1}} = \mathrm{m_2B^{c_2}}$$

$$\frac{\mathrm{m_1}}{\mathrm{m_2}}\mathrm{B^{c_1}} = \mathrm{B^{c_2}}$$

(W) $$\frac{\mathrm{m_1}}{\mathrm{m_2}} = \mathrm{B^{c_2-c_1}}$$

Teraz rozważmy przypadki (W):
$1)\mathrm{c_1} = \mathrm{c_2} \rightarrow \frac{\mathrm{m_1}}{\mathrm{m_2}} = 1 \rightarrow \mathrm{m_1} = \mathrm{m_2}$ sprzeczność z (Z)
$2)\mathrm{c_1} > \mathrm{c_2} \rightarrow \mathrm{B^{c_2-c_1}} \leq \mathrm{B^{-1}} \rightarrow \frac{\mathrm{m_1}}{\mathrm{m_2}} \leq \frac{1}{B} \rightarrow \mathrm{m_1} \leq \frac{\mathrm{m_2}}{B}$
Pamiętając, że $\mathrm{m_1} \in \left[\frac{1}{\mathrm{B}}, 1\right)$ otrzymujemy $\frac{1}{B} \leq \mathrm{m_1} \leq \frac{\mathrm{m_2}}{B}$ czyli
$\frac{1}{B} \leq \frac{\mathrm{m_2}}{B}\ zatem\ m_2 \geq 1$ sprzeczność z założeniem zadania
$3)\mathrm{c_1} < \mathrm{c_2} \rightarrow \mathrm{B^{c_2-c_1}} \geq \mathrm{B} \rightarrow \frac{\mathrm{m_1}}{\mathrm{m_2}} \geq \mathrm{B} \rightarrow \mathrm{m_1} \geq Bm_2$
Pamiętając, że $\mathrm{m_2} \in \left[\frac{1}{\mathrm{B}}, 1\right)$ otrzymujemy $\mathrm{m_1} \geq Bm_2 \geq B * \frac{1}{B}$ czyli
$\mathrm{m_1} \geq 1$ sprzeczność z założeniem zadania

Stąd każda niezerowa liczba rzeczywista x ma jednoznaczne przedstawienie w postaci $\mathrm{x} = \mathrm{smB^c}$ ■

Zadanie 3. 1 punkt (komputer)

1) Jaki jest najmniejszy przedział [A, B], zawierający te liczby?

$$|\min(x)| = (0.01000)_2 = 0.25$$

$$|\max(x)| = (1.1110)_2 = 1.875$$

Zatem przedział ten jest postaci $[A, B] = [-1.875, 1.875]$
2) Jak liczby rozkładają się w [A, B]?
Wykonaj odpowiedni rysunek. Co z niego wynika?
W przedziale $[-0.25, 0.25]$ nie występują żadne liczby, w pozostałej części $[A, B]$ im bliżej górnych krańców, tym mniej gęsto rozkładają się liczby.

Zadanie 4. 1 punkt
Przeczytaj tekst dostępny pod adresem
http://www-users.math.umn.edu/ ~arnold//disasters/patriot.html
mówiący o tym, że niefrasobliwe używanie arytmetyki zmiennopozycyjnej może prowadzić do prawdziwej tragedii (szczegóły patrz raport GAO/IMTEC-92-26). Streść, własnymi słowami, opisane tam zdarzenie i przedstaw istotę opisanego problemu.

25 lutego 1991 roku, z powodu błędów numerycznych, amerykański pocisk antyrakietowy nie trafił w Iracką rakietę, która zabiła 28 osób i raniła 100 kolejnych. Błąd dotyczył niedokładnej operacji na systemowym zegarze.
Żeby otrzymać z niego liczbę sekund mnożono liczbę dziesiętnych sekundy przez 1/10. Problem w tym, że 1/10 nie ma skończonego rozwinięcia dwójkowego, a maszyny wykorzystywane do obliczeń były 24-bitowe, więc wyniki wychodziły niepoprawnie zaokrąglone.
Zamiast zaokrąglić w poprawny sposób, obcinano część liczby niemieszczącą się na tych 24-bitach, co powodowało kumulowanie się błędów.
W trakcie każdej sekundy błąd przybliżenia wynosił $9{,}5 * 10^{-8}$, więc po 100 godzinach nieustannej pracy maszyny błąd ten skumulował się do $0{,}34$ sekundy. Ponieważ wspomniana rakieta poruszała się z prędkością $1\ 676\ m/s$ a błąd skumulował się do $0{,}34$ sekundy, to w trakcie podanego czasu pokonywała ona około 570 metrów, co uniemożliwiło jej precyzyjne namierzenie i unieszkodliwienie.

Zadanie 5. 1 punkt
Zapoznaj się ze standardem IEEE 754 reprezentacji liczb zmiennopozycyjnych. Omów go krótko i podaj główne różnice w stosunku do modelu teoretycznego reprezentacji liczb maszynowych przedstawionego na wykładzie.

W standardzie IEE 754 mamy sztywno określoną ilość bitów przeznaczoną na znak (1), cechę (8 dla single, 11 dla double) i mantysę (23 single, 52 double).
Wykładnik kodowany jest z nadmiarem (127 single, 1023 double) co oznacza, że potęga 2 równa 1 jest zapisywana jako 128. Stąd zakres cechy to $[2^{-127}, 2^{128}]$ dla single oraz $[2^{-1023}, 2^{1024}]$ dla double.
Różnica między wersją modelu przedstawieniowego z wykładu a standardem IEE 754 jest zakres mantysy – w 1 przypadku to $\left[\frac{1}{2}, 1\right)$, a w drugim to $[1{,}2]$.
Z racji tego, że pierwsza cyfra mantysy w IEE 754 to zawsze 1, to pomijamy ją w zapisie mantysy, co oszczędza nam dodatkowy bit.
Standard IEE 754 zawiera specjalne wartości:
$+0, -0, +\infty, -\infty, NaN$ (nie-liczba, czasem ma 2 warianty – cichy i głośny).

Zadanie 7. 1 punkt
Wytłumacz dokładnie kiedy występuje i na czym polega
zjawisko utraty cyfr znaczących wyniku.

Utrata cyfr znaczących występuje m. in. podczas odejmowania liczb prawie równych, których różnica jest znacznie mniejsza niż każda z tych liczb.

Ogólne wyjaśnienie zjawiska:

W efekcie odejmowania 2 liczb prawie równych dochodzi do nieakceptowalnych zaokrągleń przez co liczba cyfr znaczących wyniku maleje.

Szczegółowe wyjaśnienie zjawiska:

Niech x, y będą prawie równymi liczbami rzeczywistymi, a $rd(x), rd(y)$ ich reprezentacją komputerową postaci $(*)$ $rd(z) = sm_t * 2^c$
gdzie s jest bitem znaku, m_t oznacza t-bitową mantysę taką, że $m \in \left(0; \frac{1}{2}\right]$,
natomiast c jest cechą – dowolną liczbą całkowitą.

Niech $\begin{cases} rd(x) = s\,0\,.\,1\,x_1\,\cdots\,x_i\,x_{i+1}\,\cdots\,x_t \\ rd(y) = s\,0\,.\,1\,y_1\,\cdots\,y_i\,y_{i+1}\,\cdots\,y_t \end{cases}$

Czyli niech na i pierwszych bitach $rd(x) = rd(y)$
Wtedy różnica liczb wynosi

$$R = rd(x) - rd(y) = s\,0\,.\,0\,0\,\cdots\,0\,1\,\cdots\,x_t - y_t$$

Jak widać, na i pierwszych bitach otrzymujemy same zera, na $i + 1$ jest jedynka, a na pozostałych różnica $rd(x_i) - rd(y_i)$. Nie jest to poprawna reprezentacja komputerowa liczby R, więc musimy doprowadzić ją do postaci (*) dokonując przesunięcie i-bitowe w lewo (czyli przemnażając R przez 2^i):

$$R = rd(x) - rd(y) = s\,0\,.\,1\,\cdots\,x_t - y_t\,??\,\cdots ?$$

Z racji tego, że ilość bitów przeznaczonych na reprezentację R się nie zmieniła, otrzymujemy liczbę, której i ostatnich bitów nie wpływa na wynik, stąd nazwa zjawiska – utraciliśmy i cyfr znaczących wyniku.