

Přesnost a rychlost výpočtu

Jan Faigl

Katedra počítačů
Fakulta elektrotechnická
České vysoké učení technické v Praze

Přednáška 13

BAB36PRGA – Programování v C

Přehled témat

- Část 1 – Přesnost výpočtu
Přesnost výpočtů a numerická stability
- Část 2 – Rychlost výpočtu (programu)
Maticové násobení
Rychlost výpočtu
Comparing C to Machine Code
Paralelní výpočet
- Část 3 – Implementace domácích úkolů
Diskutovaná témata

Přesnost výpočtů

Část I

Část 1 – Přesnost výpočtu

Přesnost výpočtů

Přesnost výpočtu - Příklad součtu dvou čísel

```
1 #include <stdio.h>
2
3 int main(void)
4 {
5     double a = 1e+10;
6     double b = 1e-10;
7
8     printf("a : %24.121f\n", a);
9     printf("b : %24.121f\n", b);
10    printf("a+b: %24.121f\n", a + b);
11
12    return 0;
13 }
14
15 clang sum.c && ./a.out
16 a : 10000000000.000000000000
17 b : 0.000000000000100
18 a+b: 10000000000.000000000000
```

lec13/sum.c

Přesnost výpočtů

Přesnost výpočtu - Příklad dělení dvou čísel

```
1 #include <stdio.h>
2
3 int main(void)
4 {
5     const int number = 100;
6     double dV = 0.0;
7     float fV = 0.0f;
8
9     for (int i = 0; i < number; ++i) {
10        dV += 1.0 / 10.0;
11        fV += 1.0 / 10.0;
12    }
13
14    printf("double value: %lf ", dV);
15    printf("float value: %lf ", fV);
16
17    return 0;
18 }
19
20 clang division.c && ./a.out
21 double value: 10.000000 float value: 10.000000
```

lec13/division.c

Přesnost výpočtů

Přesnost výpočtu - strojová přesnost

- Strojová přesnost ϵ_m - nejmenší desetinné číslo, které přičtením k 1.0 dává výsledek různý od 1, pro $|v| < \epsilon_m$, platí

$$v + 1.0 == 1.0.$$

Symbol == odpovídá porovnání dvou hodnot (test na ekvivalenci).

- Zaokrouhlovací chyba - nejméně ϵ_m .
- Přesnost výpočtu - aditivní chyba roste s počtem operací v řádu $\sqrt{N} \cdot \epsilon_m$.
 - Často se však kumuluje preferabilně v jedno směru v řádu $N \cdot \epsilon_m$.

Přesnost výpočtů

Zdroje a typy chyby

- Chyby matematického modelu - matematická aproximace fyzikální situace.
- Chyby vstupních dat.
- Chyby numerické metody.
- Chyby zaokrouhlovací.

- Absolutní chyba aproximace
 $E(x) = \hat{x} - x$, \hat{x} přesná hodnota, x aproximace.
- Relativní chyba $RE(x) = \frac{\hat{x} - x}{x}$.

Přesnost výpočtů

Podmíněnost numerických úloh

- Podmíněnost úlohy $C_p = \frac{\text{relativní chyba výstupních údajů}}{\text{relativní chyba vstupních údajů}}$.
- Dobře podmíněná úloha $C_p \approx 1$.
- Výpočet je dobře podmíněný, je-li málo citlivý na poruchy ve vstupních datech.
- Numericky stabilní výpočet - vliv zaokrouhlovacích chyb na výsledek je malý.
- Výpočet je stabilní, je-li dobře podmíněný a numericky stabilní.

Přesnost výpočtů

Možnosti zvýšení přesnosti

- Reprezentace racionálních čísel - podíl dvou celočíselných hodnot, např. *Homogenní souřadnice*.
- „Libovolná přesnost“ - speciální knihovny, např. *gmp* až do výše volné paměti.
<https://gmplib.org/manual/index>
 - Souřadnice x,y - 7511164176768 346868669952 3739567104 ~ 2008,57; 92,76.

Přenos výpočtů

Číslení mnoha malých necelých čísel - 1/2

- Na příkladu součtu dvou velmi odlišných čísel (např. $1 \times 10^{10} + 1 \times 10^{-10}$) dochází z důvodu omezené reprezentace mantisy k zaokrouhlovací chybě.
- V případě naivní implementace součtu velkého počtu (např. 2^{30}) velmi malých hodnot (např. 1×10^{-20}) může dojít vlivem zaokrouhlování k významné chybě.

lec13/addition.c

```

// small value to be sum
float v = 1e-20f; //float literal
// 1073741824 is 2^30 values (1e9)
const size_t power = 30;
size_t n = 1l<<power;

// multiplication factor for print
const double k = 1e11;

float sum_naive(size_t n, float *v)
{
    float r = 0;
    for (size_t i = 0; i < n; ++i) {
        r += v[i];
    }
    return r;
}

float sum_alter(size_t n, float *v, size_t power)
{
    float r = 0;
    const size_t order = power - 1;
    size_t k = 2;
    for (size_t l = 1; l < order; ++l, k *= 2) {
        for (size_t i = 0; i < n; i += k) {
            v[i] = v[i] + v[i+k/2];
        }
        k /= 2;
    }
    for (size_t i = 0; i < n; i += k) {
        r += v[i];
    }
    return r;
}

double sum1 = v*n * k;
double sum2 = sum_naive(n, values) * k;
float sum3 = sum_alter(n, values, power) * k;

Přímé násobení - výsledek      Naivní součet - výsledek      Číslení po dvojicích - výsledek
1.073 741 789 925 364 287 228 1.      0.022 737 367 544 323 205 947 9.      1.073 741 793 632 507 324 218 8.

```

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 20 / 38

Přenos výpočtů

Číslení mnoha malých necelých čísel - 2/2

```

1 #include <stdio.h>
2 #include <stdlib.h>
3
4 float* init_values(size_t n, float v);
5 float sum_naive(size_t n, float *v);
6 float sum_alter(size_t n, float *v, size_t power);
7
8 int main(void)
9 {
10     float v = 1e-20f; // small value to be sum
11     const size_t power = 30; // try 3 vs. 30
12     size_t n = 1l<<power; // 1073741824 is 2^30 values
13     const double k = 1e11;
14
15     float *values = init_values(n, v);
16
17     double sum1 = v * n * k;
18     double sum2 = sum_naive(n, values) * k;
19     float sum3 = sum_alter(n, values, power) * k;
20
21     printf("Sum of %lu numbers of the value %.22f\n", n, v);
22     printf("Sum1 (multiplication): %.22f\n", sum1);
23     printf("Sum2 (naive) : %.22f\n", sum2);
24     printf("Sum3 (alter) : %.22f\n", sum3);
25     free(values);
26     return EXIT_SUCCESS;
27 }

```

\$ clang addition.c -o addition && ./addition
Sum of 1073741824 numbers of the value 0.0000000000000000000100
Sum1 (multiplication): 1.0737417899253642872281
Sum2 (naive) : 0.0227373675443232059479
Sum3 (alter) : 1.0737417936325073242188

\$ calc "1e-20 * 2^30 * 1e11"
1.073741824

lec13/addition.c

- Implementujte s využitím knihovny `gmp`.

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 21 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Část II

Část 2 – Rychlost výpočtu (programu)

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 22 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Maticové násobení - Naivně

```

1 void simple_multiply(const int n, const double *a, const double *b, double *c)
2 {
3     for (int i = 0; i < n; ++i) {
4         for (int j = 0; j < n; ++j) {
5             double prod = 0;
6             for (int k = 0; k < n; ++k) {
7                 prod += a[i * n + k] * b[k * n + j];
8             }
9             c[i * n + j] = prod;
10        }
11    }
12 }

```

- Pro přehlednost předpokládáme kompatibilní rozměry matic a správně alokované.

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 24 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Maticové násobení - Naivně s transpozicí

```

1 void simple_multiply_trans(const int n, const double *a, const double *b, double *c)
2 {
3     double *bT = create_matrix(n); // allocate memory for transposed matrix
4     for (int i = 0; i < n; ++i) {
5         bT[i*n + i] = b[i*n + i];
6         for (int j = i + 1; j < n; ++j) {
7             bT[i*n + j] = b[j*n + i];
8             bT[j*n + i] = b[i*n + j];
9         }
10    }
11    for (int i = 0; i < n; ++i) {
12        for (int j = 0; j < n; ++j) {
13            double tmp = 0;
14            for (int k = 0; k < n; ++k) {
15                tmp += a[i*n + k] * bT[j*n + k];
16            }
17            c[i*n + j] = tmp;
18        }
19    }
20    free(bT);
21 }

```

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 25 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Porovnání rychlosti násobení matic

Maticové násobení

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 26 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Architektura procesoru a způsob výpočtu

- Příklad násobení matic a násobení transponované matice ukazuje, že kromě instrukcí má zásadní vliv **organizace dat a přístup do paměti**.
- V moderních procesorech hraje **cache** zásadní roli společně s řetězením instrukcí, tzv. **pipelining**, a využitím specifických instrukcí.
- SIMD - Single Instruction Multiple Data
- Proniknutí do detailů fungování cache a řetězení instrukcí je náplní předmětu **Architektura počítačů (BOB35APO)**, kde máte možnost se seznámit s přicházející architekturou RISC V.

<https://cw.felk.cvut.cz/wiki/courses/b35apo/>

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 28 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Optimalizace kódu

- Kromě optimalizace výsledného spustitelného kódu při prekládání, je možné optimalizovat kódu za běhu nebo již existujících binární (přeložené) soubory.
 - BOLT** - Binary Optimization and Layout Tool, zrychlení o až 20%–50%
<https://arxiv.org/abs/1807.06735>
<https://dl.acm.org/doi/10.5555/3314872.3314876>
- Využití speciálních instrukcí v základních funkcích může výpočty (programy) výrazně urychlit, zejména pokud se funkce používají masivně.
 - AVX2 a EVEX instrukce (ze sady SSE4.2) ve funkcích porovnání řetězců `str{n}.casecmp()` – až o 38% méně potřebného času.
03/2022 - <https://www.phoronix.com/news/Glibc-strcasecmp-AVX2-EVEX>
- V obou případech (a obecně) je vhodné rozumět principu a využít instrukce Assembleru.*

Informativní

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 29 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Compiler Explorer

<https://godbolt.org/z/K9r1eWqcd>

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přenos a rychlost výpočtu 30 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Compiler Explorer – Analýza optimalizovaného kódu

- Vliv optimalizace -O2 na výsledný kód, který obsahuje nedefinované chování, přetečení celého císla. Příloha 3. přednášky.

<https://godbolt.org/z/G3GEz4vbv>

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přesnost a rychlost výpočtu 31 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Comparing C to Machine Code

<https://www.youtube.com/watch?v=y0yaJXpAYZQ>

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přesnost a rychlost výpočtu 33 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Příklad použití OpenMP - Maticové násobení 1/2

- Open Multi-Processing (OpenMP) - aplikační programové rozhraní (API) multiplatformních výpočtů se sdílenou pamětí. <http://www.openmp.org>
- Direktivou preprocesoru můžeme instruovat kompilátor k vytvoření kódu paralelního výpočtu, např. paralelizace přes větší proměnnou `i`.

```

1 void multiply(int n, int a[n][n], int b[n][n], int c[n][n])
2 {
3     int i;
4     #pragma omp parallel private(i)
5     #pragma omp for schedule (dynamic, 1)
6     for (i = 0; i < n; ++i) {
7         for (int j = 0; j < n; ++j) {
8             c[i][j] = 0;
9             for (int k = 0; k < n; ++k) {
10                c[i][j] += a[i][k] * b[k][j];
11            }
12        }
13    }
14 }

```

`lec13/demo-omp-matrix.c`

Pro přehlednost uvažujeme čtvercové matice stejných rozměrů.

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přesnost a rychlost výpočtu 35 / 38

Maticové násobení Rychlost výpočtu Comparing C to Machine Code Paralelní výpočet

Příklad použití OpenMP - Maticové násobení 2/2

- Příklad násobení matic 1000×1000 s využitím OpenMP na iCore5 (2 jádra s HT ~ 4x výpočetní jednotky).
- Násobení matic 5000×5000 (Ryzen 9 5950X).

```

1 gcc -std=c99 -O2 -o demo-omp demo-omp-matrix.c -fopenmp
2 ./demo-omp 1000
3 Size of matrices 1000 x 1000 naive
4 multiplication with 0(n^3)
5 c1 == c2: 1
6 Multiplication single core 9.33 sec
7 Multiplication multi-core 4.73 sec
8
9 OMP_NUM_THREADS=2 ./demo-omp 1000
10 Size of matrices 1000 x 1000 naive
11 multiplication with 0(n^3)
12 c1 == c2: 1
13 Multiplication single core 9.48 sec
14 Multiplication multi-core 6.23 sec

```

```

1 OMP_NUM_THREADS=16 ./demo-omp 5000
2 Size of matrices 5000 x 5000 naive
3 multiplication with 0(n^3)
4 Multiplication single core 256.00 sec
5 Multiplication multi-core 18.05 sec

```

`lec13/demo-omp-matrix.c`

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přesnost a rychlost výpočtu 36 / 38

Diskutovaná témata

Shrnutí přednášky

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přesnost a rychlost výpočtu 37 / 38

Diskutovaná témata

- Numerická přesnost.
- Knihovna gmp.
- Maticové násobení a organizace paměti.
- Rychlost výpočtu a architektura procesoru.
- Paralelní výpočty OpenMP.
- Domácí úkol HW10B.

Jan Faigl, 2024 BAB36PRGA – Přednáška 13: Přesnost a rychlost výpočtu 38 / 38