Minimální reprezentace víceintervalových booleovských funkcí

Obhajoba diplomové práce

Filip Bártek

Katedra teoretické informatiky a matematické logiky

4. června 2015

Motivace

- Generování testovacích vzorů (hardware, software)
- Omezující podmínky na celočíselných proměnných
  - Lineární (nerovnosti) $\rightarrow$ intervaly
  - Nelineární $\rightarrow$ DNF

1-intervalové funkce

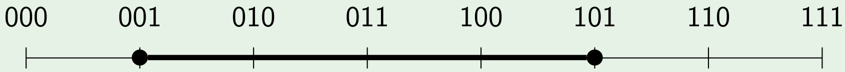
n-bitové celé číslo $\sim$ binární vektor délky n

Definice (1-intervalová funkce [Schieber et al., 2005])

$f_{[a,b]}(x) = 1 \iff a \leq x \leq b$

Příklad (1-intervalová funkce)

$f_{[001,101]}$



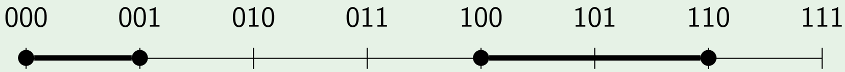
k-intervalové funkce

Definice (k-intervalová funkce)

$f_{[a_1,b_1],\ldots,[a_k,b_k]}(x) = 1 \iff a_i \leq x \leq b_i$ pro některé i

Příklad (2-intervalová funkce)

$f_{[000,001],[100,110]}$

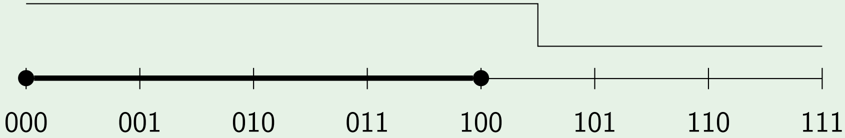


l-zlomové funkce

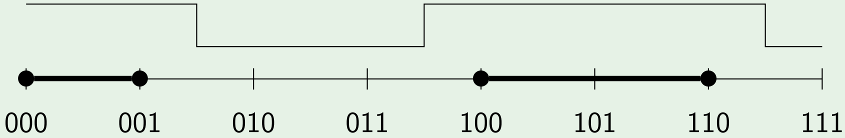
Definice (l-zlomová funkce [Hušek, 2014])

Funkce f je l-zlomová, pokud existuje právě l vektorů x takových, že $f(x) \neq f(x+1)$.

Příklad (1-zlomová funkce)



Příklad (3-zlomová funkce)



Disjunktivní normální forma a ternární vektory

term $\sim$ *ternární vektor*, tj. vektor nad abecedou $\{0, 1, \phi\}$

Příklad (Term a ternární vektor)

$$x_1\overline{x_4} \sim 1\phi\phi0$$

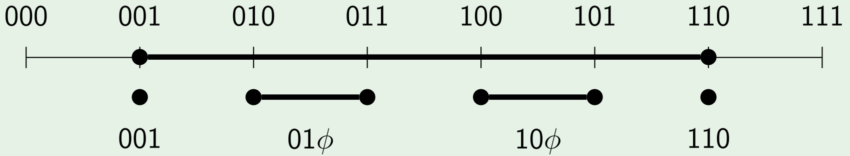
Pokrytí

Definice (Pokrytí)

Pokrytí funkce f je množina ternárních vektorů ekvivalentní DNF reprezentaci f.

Příklad (Pokrytí)

$$f_{[001,110]} \sim \{001, 01\phi, 10\phi, 110\}$$



Minimalizace pokrytí k-intervalové funkce

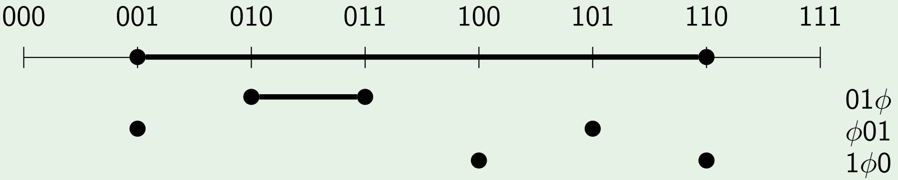
Problém

Vstup *n-bitová čísla $a_1, b_1, \ldots, a_k, b_k$*

Výstup *Minimální pokrytí k-intervalové funkce $f_{[a_1,b_1],\ldots,[a_k,b_k]}$*

Příklad (Minimální pokrytí 2-intervalové funkce)

$$f_{[001,110]} \sim \{01\phi, \phi01, 1\phi0\}$$



Teoretická východiska

- Schieber et al. [2005]
  - Optimalizační algoritmus pro 1-intervalové funkce
- Dubovský [2012]
  - Optimalizační algoritmus pro 2-zlomové 2-intervalové funkce
  - 2-aproximační algoritmus pro 2-intervalové funkce
  - 3-zlomová a 4-zlomová funkce, které nejsou coverable

Nové výsledky v diplomové práci

- Zjednodušený optimalizační algoritmus pro 2-zlomové 2-intervalové funkce

Nové výsledky v diplomové práci

- Zjednodušený optimalizační algoritmus pro 2-zlomové 2-intervalové funkce
- Technika použitá v důkazech optimality pokrytí 2-zlomových funkcí není obecně použitelná pro 3- a vícezlomové funkce

Nové výsledky v diplomové práci

- Zjednodušený optimalizační algoritmus pro 2-zlomové 2-intervalové funkce
- Technika použitá v důkazech optimality pokrytí 2-zlomových funkcí není obecně použitelná pro 3- a vícezlomové funkce
- $2k$-aproximační algoritmus pro k-intervalové funkce pro každé $k \geq 0$

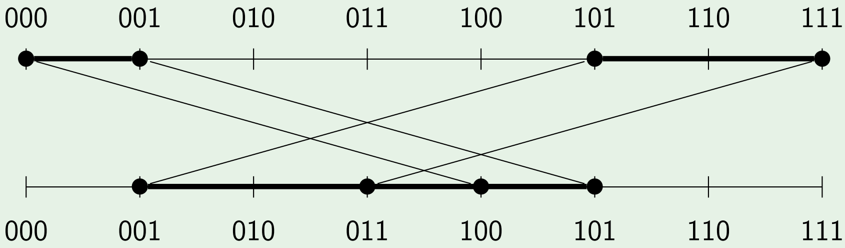
Nové výsledky v diplomové práci

- Zjednodušený optimalizační algoritmus pro 2-zlomové 2-intervalové funkce
- Technika použitá v důkazech optimality pokrytí 2-zlomových funkcí není obecně použitelná pro 3- a vícezlomové funkce
- $2k$-aproximační algoritmus pro k-intervalové funkce pro každé $k \geq 0$
- Přirozené vylepšení aproximačního algoritmu a důkaz, že není o moc lepší

Zjednodušený algoritmus pro 2-zlomové funkce

2-zlomová 2-intervalová funkce $\mapsto$ 1-intervalová funkce

Příklad (Transformace 2-zlomové funkce na 1-intervalovou)



Non-coverable l-zlomová funkce pro každé $l \geq 3$

Dubovský [2012]: 3-zlomová funkce $f_{[0,4],[9,14]}$ není coverable.
Bártek [2015]: l-zlomová non-coverable funkce pro každé $l \geq 3$

Závěr

Optimalitu algoritmu pro třídu všech k-intervalových funkcí pro $k \geq 2$ nelze dokázat konstrukcí ortogonální množiny.

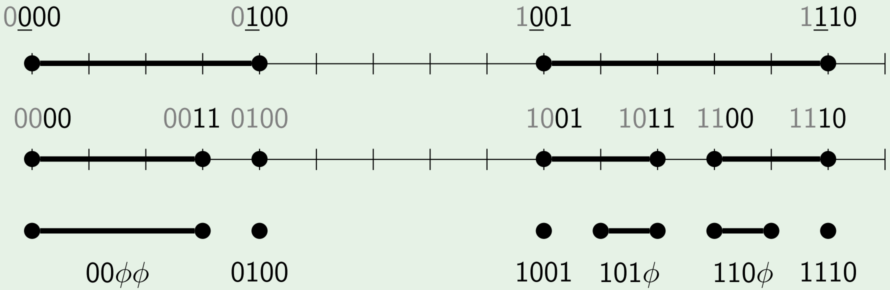
Jednoduchý $2k$-aproximační algoritmus

Algoritmus (Suffix-prefix dekompozice)

Každý interval rozdělíme na dva podintervaly (suffixový a prefixový).
Každý z těchto podintervalů pokryjeme zvlášť optimálně.

Aproximační poměr: $2k$

Příklad (Suffix-prefix dekompozice)



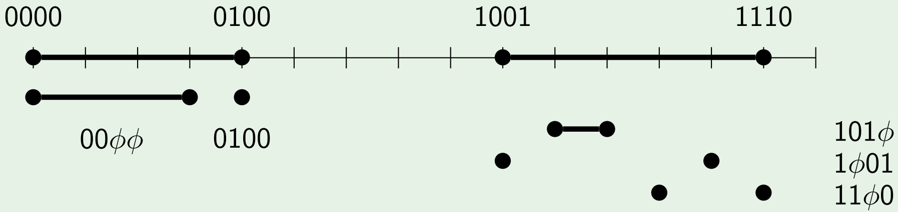
Vylepšený $2k$-aproximační algoritmus

Algoritmus (Intervalová dekompozice)

Každý interval pokryjeme zvlášť optimálně.

Aproximační poměr: mezi $2k - 2$ a $2k$

Příklad (Intervalová dekompozice)



Shrnutí

- Dosud používaná technika důkazu optimality pokrytí selhává u 3- a vícezlomových funkcí.

Shrnutí

- Dosud používaná technika důkazu optimality pokrytí selhává u 3- a vícezlomových funkcí.
- Ortogonální množiny však stále mohou dávat horní odhad aproximačního poměru.

Shrnutí

- Dosud používaná technika důkazu optimality pokrytí selhává u 3- a vícezlomových funkcí.
- Ortogonální množiny však stále mohou dávat horní odhad aproximačního poměru.
- Dekompozice na intervaly se pro velké k nechová o mnoho lépe než $2k$-aproximačně.

Shrnutí

- Dosud používaná technika důkazu optimality pokrytí selhává u 3- a vícezlomových funkcí.
- Ortogonální množiny však stále mohou dávat horní odhad aproximačního poměru.
- Dekompozice na intervaly se pro velké k nechová o mnoho lépe než $2k$-aproximačně.

- Do budoucna:
  - Efektivní optimalizační algoritmus pro třídu všech k-intervalových funkcí pro nějaké $k \geq 2$
  - Aproximační algoritmus pro obecné k s aproximačním poměrem lepším než $2k - 2$

Shrnutí

- Dosud používaná technika důkazu optimality pokrytí selhává u 3- a vícezlomových funkcí.
- Ortogonální množiny však stále mohou dávat horní odhad aproximačního poměru.
- Dekompozice na intervaly se pro velké k nechová o mnoho lépe než $2k$-aproximačně.

- Do budoucna:
  - Efektivní optimalizační algoritmus pro třídu všech k-intervalových funkcí pro nějaké $k \geq 2$
  - Aproximační algoritmus pro obecné k s aproximačním poměrem lepším než $2k - 2$

Děkuji za Vaši pozornost.

Záložní slidy

Výpočetní složitost algoritmů v DP

Výpočetní složitost všech uvedených algoritmů je *polynomiální* vzhledem k n i k.

- Procedury:
  - Prefix: $T_{prefix}(n) \in O(n^2)$
  - 1 interval: $T_{INT(1)}(n) \in O(n^3)$
    - ★ Algoritmus redukuje na jednu menší instanci + polynomiální processing:

$$T(n) \leq T(n-1) + O(n^2) \leq nO(n^2) + O(1) \in O(n^3)$$

- Vlastní algoritmy:
  - 2-zlomové 2-intervalové funkce: $T_{SWITCH(2)}(n) \in O(n^3)$
  - Suffix-prefix dekompozice: $T_{SPD}(n,k) \in O(kT_{prefix}(n)) = O(kn^2)$
  - Intervalová dekompozice: $T_{ID}(n,k) \in O(kT_{INT(1)}(n)) = O(kn^3)$

Permutace souřadnic

Permutaci souřadnic jsem zvážil pouze jako pre-processing *mezí*, kde obecně nefunguje (nezachovává pokrytí).

Příklad (Permutace souřadnic v mezích)

$a = 001, b = 110, \mathcal{T}_{[a,b]} = \{01\phi, \phi01, 1\phi0\}$
$\pi = (2,3)$
$\pi(a) = 010, \pi(b) = 101, \mathcal{T}_{[\pi(a),\pi(b)]} = \{01\phi, 10\phi\}, \pi^{-1}(\mathcal{T}_{[\pi(a),\pi(b)]}) = \{0\phi1, 1\phi0\}$

Permutace souřadnic by šlo použít i pro post-processing výstupu aproximačního algoritmu a je to zajímavý nápad otevřený dalšímu zkoumání.

Suffix-prefix dekompozice – zdůvodnění

Proč jsem nezobecnil přímo algoritmus z Schieber et al. [2005]?

- Jednodušší analýza aproximačního poměru
- Aproximační poměr je nezměněn ($2k$)
- Algoritmus *intervalová dekompozice* (uvedený v následující kapitole) je lepší
- Algoritmus *suffix-prefix dekompozice* především dává horní odhad $2k$ na aproximační poměr algoritmu *intervalová dekompozice*

Suffix-prefix dekompozice – zlepšený algoritmus

Algoritmus (Zlepšená suffix-prefix dekompozice)

Každý interval pokryjeme zvlášť. *Pokud je interval po odstranění společného prefixu mezí prefixový* ($a = 0^{\{m\}}$) *nebo suffixový* ($b = 1^{\{m\}}$)*, pokryjeme jej optimálně.* Jinak jej rozdělíme na suffixový a prefixový interval a každý z nich pokryjeme optimálně.

Dolní odhad aproximačního poměru $2k$ dostaneme pomocí množiny „zlých“ funkcí:

$$f_l^n \sim \{[p0^{\{n-l-1\}}1, p1^{\{n-l-1\}}0] | p \in \{0,1\}^l\}$$

Počet intervalů f_l^n: $k = 2^l$
Velikost optima: $n - l$
Velikost výstupu algoritmu: $2k(n-l) - 2k$

Závěr

Proti každému aproximačnímu poměru menšímu než $2k$ (pro $k = 2^l$) existuje protipříklad.

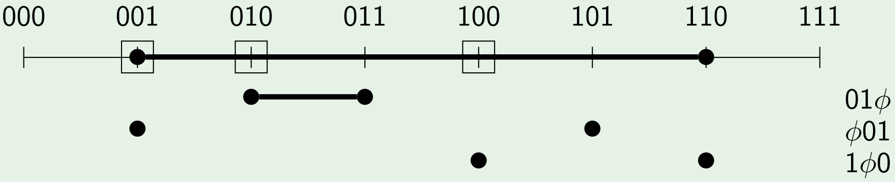
Ortogonální množina

Definice (Ortogonální množina)

Ortogonální množina funkce f je množina true pointů f takových, že žádné dva z nich nelze pokrýt jedním ternárním vektorem.

Příklad (Ortogonální množina)

$$V = \{001, 010, 100\}$$



Coverable funkce

Definice (Coverability)

Funkce je *coverable*, pokud má pokrytí a ortogonální množinu stejné velikosti.

Pozorování

Takové pokrytí je nutně minimální.

Bibliografie

Filip Bártek. Minimální reprezentace víceintervalových booleovských funkcí. Master's thesis, Charles University in Prague, 2015.

Richard A. DeMillo and A. Jefferson Offutt. Constraint-based automatic test data generation. *Software Engineering, IEEE Transactions on*, 17(9):900–910, Sep 1991. doi: 10.1109/32.92910. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=92910.

Jakub Dubovský. A construction of minimum DNF representations of 2-interval functions. Master's thesis, Charles University in Prague, 2012. URL https://is.cuni.cz/webapps/zzp/detail/116613/.

Radek Hušek. Properties of interval Boolean functions. Master's thesis, Charles University in Prague, 2014. URL https://is.cuni.cz/webapps/zzp/detail/145114/. [In Czech].

Daniel Lewin, Laurent Fournier, Moshe Levinger, Evgeny Roytman, and Gil Shurek. Constraint satisfaction for test program generation. In *Computers and Communications, 1995., Conference Proceedings of the 1995 IEEE Fourteenth Annual International Phoenix Conference on*, pages 45–48. IEEE, Mar 1995. doi: 10.1109/PCCC.1995.472513. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=472513.

Baruch Schieber, Daniel Geist, and Ayal Zaks. Computing the minimum DNF representation of Boolean functions defined by intervals. *Discrete Applied Mathematics*, 149(1–3):154 – 173, 2005. ISSN 0166-218X. doi: http://dx.doi.org/10.1016/j.dam.2004.08.009. URL http://www.sciencedirect.com/science/article/pii/S0166218X05000752.