On the efficiency of normal form systems of Boolean functions

EJCIM – Student presentations

Pierre Mercuriali
Joint work with
Miguel Couceiro, Erkko Lehtonen, Romain Péchoux, Mathias Soeken

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

29 of March, 2018

Plan

1. Preliminaries:
   - Boolean functions,
   - Clones,
   - Normal Form Systems (**NFS**s)
2. Efficiency of NFSs
   - How to measure efficiency?
   - Classification of NFSs
3. Future work

Context

- Representation of Boolean functions
- Efficient representations? Number of connectives
- Here: stratified formulas (connectives occur in constrained order)
  Variants: Jukna, 2012
- Median Normal Form: shown to be "more efficient" than DNF, CNF, etc.
- Other connectives/ Normal Form Systems?

Clones of Boolean functions

Class composition of K with J:

$$K \circ J = \{f(g_1, \ldots, g_n) \colon f \text{ } n\text{-ary in } K,\, g_1, \ldots, g_n \text{ } m\text{-ary in } J\}$$

Definition

A **clone** is a class $C \subseteq \Omega$ that contains all projections and satisfies $C \circ C = C$.

Examples of clones:

- Clone of all projections: I_c
- Clone of literals and constants: $\Omega(1)$
- Clone of all conjunctions: Λ
- Clone of all monotone functions: M
- Clone of all Boolean functions: Ω

Known results about (Boolean) clones:

- Clones constitute an algebraic lattice (E. Post, 1941).
  - Largest clone: Ω
  - Smallest clone: I_c

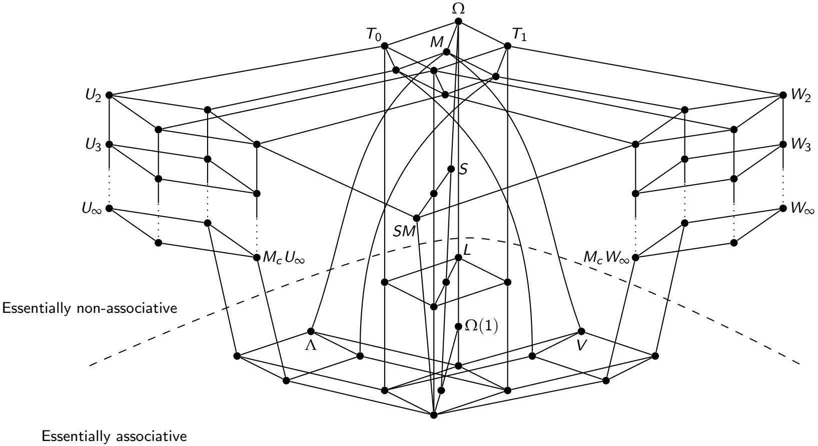
- Each clone C is finitely generated: $C = \mathcal{C}(K)$, for some finite $K \subseteq \Omega$ with:

$$\mathcal{C}(K) \quad = \quad \bigcap_{K \subset C \text{ clone}} C$$

- Each C has a dual clone $C^d = \{f^d\colon\ f \in C\}$, with

$$f^d(x_1, \ldots, x_n) = \overline{f(\overline{x_1}, \ldots, \overline{x_n})}$$

Clone essentially associative: all essential functions are associative



Essentially unary clones: generated by essentially unary functions

- $I_c = \mathcal{C}(\{\,\})$, $I_0 = \mathcal{C}(\{\mathbf{0}\})$, $I_1 = \mathcal{C}(\{\mathbf{1}\})$ and $I = \mathcal{C}(\{\mathbf{0}, \mathbf{1}\})$
- $I^* = \mathcal{C}(\{\neg\})$ **and** $\Omega(1) = \mathcal{C}(\{\mathbf{0}, \mathbf{1}, \neg\})$

Minimal clones: clones that cover the clone I_c of projections

- $\Lambda_c = \mathcal{C}(\{\wedge\})$ of conjunctions and $V_c = \mathcal{C}(\{\vee\})$ of disjunctions
- $L_c = \mathcal{C}(\{\oplus\})$ of constant-preserving linear functions
- $SM = \mathcal{C}(\mathrm{m}_3)$ of self-dual ($f = f^d$) monotone functions

Known results about composition of clones:

- $C_1 \circ C_2$ of clones is **not** always a clone: $I^* \circ \Lambda$ is not a clone
- Composition of clones completely described by Couceiro, Foldes, Lehtonen (CFL2006)
- All factorizations of Ω into a composition of "prime" clones (CFL2006)
- All factorizations of Ω into a composition of minimal clones (CFL2006)

(Descending) Irredundant Factorizations of Ω:

- **DNF**: $\Omega = V_c \circ \Lambda_c \circ I^*$
- **CNF**: $\Omega = \Lambda_c \circ V_c \circ I^*$
- **PNF**: $\Omega = L_c \circ \Lambda_c \circ I$
- **PNF^d**: $\Omega = L_c \circ V_c \circ I$
- **MNF**: $\Omega = SM \circ \Omega(1)$

Each corresponds to a **normal form system** (**NFS**)

Formalizing **NFS**s

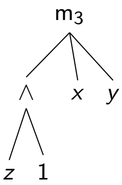
Connectives $\alpha_1, \ldots, \alpha_n$

Set of terms $T(\alpha_1 \cdots \alpha_n)$ contains:

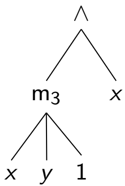
- All variables,
- All constant symbols,
- All terms $\alpha_k(t_1, \ldots, t_{ar(\alpha_k)})$ if t_i are terms

The connectives are taken in order!

In $T(\mathrm{m}_3 \wedge)$:



In $T(\wedge\, \mathrm{m}_3)$:



are not in the same NFSs!

Some **NFS**s of interest

- $\mathbf{M} = T(\mathrm{m}_3\,\neg)$ — Median NF
- $\mathbf{M}_{2n+1} = T(\mathrm{m}_{2n+1}\,\neg)$ — $2n+1$-MNF
- $\mathbf{S} = T(\uparrow) \quad (NAND)$ — Sheffer NF
- $\mathbf{S}^d = T(\downarrow) \quad (NOR)$ — Peirce NF
- $\mathbf{D} = T(\vee \wedge \neg)$ — DNF
- $\mathbf{C} = T(\wedge \vee \neg)$ — CNF
- $\mathbf{P} = T(\oplus\wedge)$ — Reed-Muller NF
- $\mathbf{P}^d = T(\oplus\vee)$ — Polynomial Dual NF

A : **NFS**, $F_{\mathbf{A}}$: set of formulas of **A**

The **A-complexity** of a Boolean function f is

$$C_{\mathbf{A}}(f) := \min\{|\phi| : \ \phi \text{ represents } f \text{ and } \phi \in F_{\mathbf{A}}\}$$

NB: Members of $\Omega(1)$ are not counted in $|\phi|$

Example:

$\mathbf{M}$: $\phi = \mathrm{m}_3(x_1, x_2, x_3)$ and $C_{\mathbf{M}}(\mathrm{MAJ}_3) = 1$

$\mathbf{D}$: $\phi = (x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_2 \wedge x_3)$ and $C_{\mathbf{D}}(\mathrm{MAJ}_3) = 5$

$\mathbf{C}$: $\phi = (x_1 \vee x_2) \wedge (x_1 \vee x_3) \wedge (x_2 \vee x_3)$ and $C_{\mathbf{C}}(\mathrm{MAJ}_3) = 5$

$\mathbf{P}$: $\phi = \oplus_3(x_1 \wedge x_2, x_1 \wedge x_3, x_2 \wedge x_3)$ and $C_{\mathbf{P}}(\mathrm{MAJ}_3) = 4$

$\mathbf{P}^d$: $\phi = \oplus_3(x_1 \vee x_2, x_1 \vee x_3, x_2 \vee x_3)$ and $C_{\mathbf{P}^d}(\mathrm{MAJ}_3) = 4$

An **NFS** **A** is polynomially as efficient as **B**, denoted $\mathbf{A} \preceq \mathbf{B}$, if there is a polynomial p with integer coefficients such that

$$C_{\mathbf{A}}(f) \leq p(C_{\mathbf{B}}(f)) \quad \text{for all } f \in \Omega$$

NB: $\preceq$ is a *quasi-ordering* of **NFS**s

If $\mathbf{A} \not\preceq \mathbf{B}$ and $\mathbf{B} \not\preceq \mathbf{A}$ holds, **then** **A** and **B** are incomparable

If $\mathbf{A} \preceq \mathbf{B}$ but $\mathbf{B} \not\preceq \mathbf{A}$, **then** **A** is polynomially more efficient than **B**

If $\mathbf{A} \preceq \mathbf{B}$ and $\mathbf{B} \preceq \mathbf{A}$, **then** **A** and **B** are equivalently efficient ($\mathbf{A} \sim \mathbf{B}$)

Motivation

Theorem (CFL2006)

1. **D**, **C**, **P**, and $\mathbf{P}^d$ are incomparable
2. **M** is polynomially more efficient than **D**, **C**, **P**, and $\mathbf{P}^d$

Definition (to be justified below)

An **NFS** **A** is *efficient* if $\mathbf{A} \sim \mathbf{M}$.

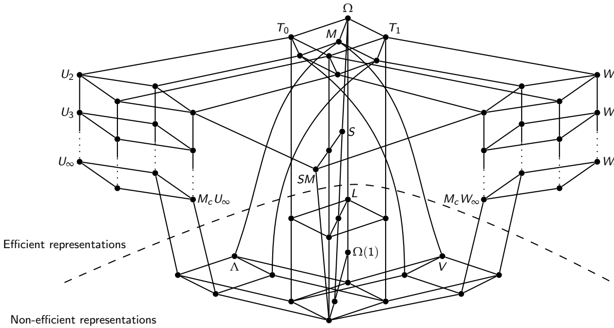
Problem 1. Existence of other **NFS**s? **E.g.:** (other connectives)

Problem 2. Classification of **NFS**s in terms of efficiency

Problem 3. Does the choice of generators within **NFS**s impact efficiency?
E.g.: m_3 vs m_5?

Problem 4. How to obtain optimal representations in each efficient **NFS**?
E.g.: optimal median normal forms?

Locating efficient **NFS**s...

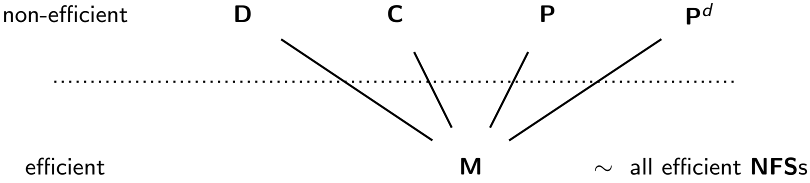


Theorem

NFSs *based on a single nontrivial connective are efficient*

Theorem

The choice of connective does not impact efficiency (ex.: $T(\mathrm{m}_3 \neg) \sim T(\mathrm{m}_5 \neg)$*)*



Theorem

M *is optimal: there is no* **NFS** *strictly below it*

NB: justifies the definition of **efficiency**!

Why is **M** optimal? (Illustration)

Property of the ternary median: pivotal function!

Definition

Median decomposition scheme (Marichal, 2009):
f a monotone Boolean function;
for any $k \in \{1, \ldots, \mathrm{ar}(f)\}$:

$$f(\mathbf{x}) = \mathrm{m}(f(\mathbf{x}_k^0), x_k, f(\mathbf{x}_k^1))$$

$\rightarrow$ Provides efficient (i.e. polynomial at most) ways to rewrite terms **A** $\rightarrow$ **M**

Exemple

Example: $f(x, y, z) = (x \wedge y) \wedge z$.

From the median decomposition scheme:

$$f(x, y, z) = \mathrm{m}(f(0, y, z), x, f(1, y, z)),$$

$$\dots$$

$$f(x, y, z) = \mathrm{m}(\mathrm{m}(\mathrm{m}(0, z, 1), y, \mathrm{m}(0, z, 0)), \underset{\uparrow}{\mathbf{x}}, \mathrm{m}(\mathrm{m}(0, z, 0), y, \mathrm{m}(0, z, 1)))$$

$\rightarrow$ Composition without (too many) repeted subterms!

Future work

1. Finer comparison of efficient NFSs
2. Redundant factorizations of Ω
3. NFSs to represent functions from a smaller clone than Ω (e.g. M)
4. Representation of multi-valued operations $\{0, \ldots, n\}^k \to \{0, \ldots, n\}$
5. Median normal forms (in **M**)
    - Decision problems: minimization, rewriting
    - Structural description

Merci de votre attention !

Thank you for your attention!

Grazie mille per la vostra attenzione!