A conjunctive logical characterization of nondeterministic linear time

Frédéric OLIVE

GREYC, Université de Caen, France
email: olive@logique.jussieu.fr

Abstract. We denote by $\mathcal{G}(\mathbf{f}, \mathbf{0}, \mathbf{s})$ the class of formulas of the form $\exists \mathbf{g} \forall \mathbf{x} \phi$, where $\mathbf{g}$ is a tuple of unary function symbols, x is a first-order variable, and ϕ quantifier-free formula of signature $\{\mathbf{g}, \mathbf{f}, \mathbf{0}, \mathbf{s}\}$.
We prove the *reduction* of $\mathcal{G}(\mathbf{f}, \mathbf{0}, \mathbf{s})$ to the subclass $\mathcal{G}^+_\wedge(\mathbf{f}, \mathbf{0}, \mathbf{s})$ of formulas written without negation nor disjunction.
As a corollary we obtain a new logical characterization of the class $\mathrm{NTIME}_{\mathrm{RAM}}(n)$: the problems of this class, and notably most of the natural NP-complete problems, can be considered as models of some sentence of the type $\exists g_1 \ldots \exists g_p \forall x \bigwedge \sigma(x) = \tau(x)$ (g_i: unary function symbols ; σ, τ: compositions of such functions), that is, in some sense, of some system of functional equations.

Introduction

Logical characterizations of complexity classes have long been studied with three goals in mind:

- establish *relationship* between *logic* and *complexity theory* in order to transfer in one of these domain some questions *a priori* related to the other. In particular, we expect this correspondance to allow proofs of nontrivial lower bound of complexity ;
- guarantee the *robustness* of complexity classes in defining them without help of any computational model;
- exhibit *complete* problems in complexity classes, to improve our understanding of all the problems they contain.

According to these three requirements, the simpler the logic, the more useful the characterization will be. This remark leads to the search for *logical reductions*: given a characterization of a complexity class $\mathcal{C}$ by a logic $\mathcal{L}$ (abusively denoted by $\mathcal{C} = \mathcal{L}$), one looks for a sublogic $\mathcal{L}' \subset \mathcal{L}$ syntaxically simpler but as expressive as $\mathcal{L}$. When such a logic $\mathcal{L}'$ exists, we say that $\mathcal{L}$ *reduces* to $\mathcal{L}'$ and that the characterization $\mathcal{C} = \mathcal{L}'$ is a *normalization* of the characterization $\mathcal{C} = \mathcal{L}$.

Here a such normalization is given for a logical characterization of nondeterministic linear time. In the 80's indeed, Grandjean has proved that the class NLIN of problems computable in linear time by a nondeterministic RAM is characterized by the class $\mathcal{G}(0, s, <)$ of formulas of type $\exists \boldsymbol{g} \forall x \phi$, where $\boldsymbol{g}$ is a tuple of unary function symbols, x is a first-order variable and ϕ is a quantifier-free formula whose signature contains $\{\boldsymbol{g}, 0, s, <\}$ (s is the successor function and $<$ is a linear order).

We show that we do not lose any expressivity in restricting $\mathcal{G}(0, s, <)$ to its subclass $\mathcal{G}^+_\wedge(0, s)$ of formulas written without $<$, without *negation*, and without *disjunction*.

Actually, this reduction holds for finite structures of the form $\langle n, \boldsymbol{f}\rangle$, where $n = \{0, \dots, n-1\}$ is an initial segment of $\mathbb{N}$ and $\boldsymbol{f}$ is a tuple of unary functions on n. Such structures being called *unary structures*, we precisely prove:

> *for each formula $\Phi \in \mathcal{G}(0, s, <)$, there exists a formula $\Phi_\wedge \in \mathcal{G}^+_\wedge(0, s)$ such that Φ and $\Phi_\wedge$ have the same unary models.*

Up to the presentation of combinatorial problems as sets of unary structures, this reduction leads to the following normalization of Grandjean's characterization of NLIN:

> *the problems belonging to NLIN (and most of the natural NP-complete problems do) are exactly the sets of unary models of formulas of the form: $\exists \boldsymbol{g} \forall x \phi$, where ϕ is a simple conjunction of equalities.*

Our plan follows:

- In Section 1, we introduce some definitions and notations.
- In Section 2, we prove a technical lemma that will highly simplify the presentation of the proof of our main result. More precisely, we state the definability in $\mathcal{G}^+_\wedge(0, s)$ of some unary arithmetical operators.
- In Section 3, we prove that every formula of $\mathcal{G}(0, s, <)$ can be considered without loss of generality as belonging to $\mathcal{G}^+(0, s)$ (*i.e.*, we prove the reduction from $\mathcal{G}(0, s, <)$ to $\mathcal{G}^+(0, s)$ over unary structures).
- In Section 4, we establish the fact that a certain kind of function defined *by cases* are definable in $\mathcal{G}^+_\wedge(0, s)$. This will be a key tool for our purpose.
- In Section 5, we easily prove our main result, that is, the reduction of $\mathcal{G}(0, s, <)$ to $\mathcal{G}^+_\wedge(0, s)$ over unary structures. The normalization of Grandjean's characterization of NLIN will then appear as an immediate corollary.

1 Preliminaries

Definition 1. Let $\mathcal{T}$ be any set of relation and function symbols.

- $\mathcal{G}(\mathcal{T})$ is the class of second-order formulas of the form $\exists \boldsymbol{g} \forall x \phi$, where $\boldsymbol{g}$ is a tuple of unary function symbols, x is a (unique) first-order variable, and ϕ is a quantifier-free formula of signature $\mathcal{T} \cup \{\boldsymbol{g}\}$.
- $\mathcal{G}^+(\mathcal{T})$ is the class of formulas belonging to $\mathcal{G}(\mathcal{T})$ and which matrix (the above "ϕ") does not contain any negation.
- $\mathcal{G}^+_\wedge(\mathcal{T})$ is the class of formulas belonging to $\mathcal{G}^+(\mathcal{T})$ and which matrix does not contain any disjunction.

Notation.

- We denote by 0 and s the predefined function symbols interpreted on every domain $n = \{0, \ldots, n-1\}$ by: for each $x < n$, $0(x) = 0$, $s(x) = x + 1$ if $x < n - 1$ and $s(n-1) = n - 1$.
- Let c be a non-negative integer. We denote by s^c, M_c, Q_c, R_c, $\mathcal{X}_c$ the predefined function symbols interpreted on every domain $n = \{0, \ldots, n-1\}$ by: for each $x < n$,

$$s^c(x) = \begin{cases} x + c & \text{if } x + c < n; \\ n - 1 & \text{else,} \end{cases}$$

$$M_c(x) = \begin{cases} cx & \text{if } cx < n; \\ n - 1 & \text{else,} \end{cases}$$

$$\mathcal{Q}_c(x) = \lfloor x/c \rfloor,$$

$$\mathcal{R}_c(x) = x \bmod c,$$

$$\mathcal{X}_c(x) = \begin{cases} 1 & \text{if } x = c; \\ 0 & \text{else.} \end{cases}$$

- For each $c \in \mathbb{N}$ we denote by $Arith_c$ the set $\{+_c, \times_c, \mathcal{Q}_c, \mathcal{R}_c, \mathcal{X}_c\}$ of the predefined symbols above defined.

Definition 2.

- A *unary signature* is a tuple of unary function symbols.
- A *unary structure* is a finite structure of domain $n \in \mathbb{N}$ over a unary signature.
- Let $\mathcal{T}$ be a unary signature and $\mathcal{T}_p$ a set of predefined symbols (typically, $\mathcal{T}_p$ can be the set $\{0, s, <\}$). Let Φ be a $(\mathcal{T} \cup \mathcal{T}_p)$-second-order formula. We denote by $\mathrm{mod}_u(\Phi)$ the set of unary structures which signature $\mathcal{T}'$ contains $\mathcal{T}$ and which expansion to $\mathcal{T}' \cup \mathcal{T}_p$ satisfies Φ.

2 Arithmetical definability in $\mathcal{G}^+_\wedge(0, s)$

Lemma 3. *Let $\mathcal{T}$ be a unary signature. Suppose that c is a predefined constant symbol definable in $\mathcal{G}^+_\wedge(0, s)$. Then for each formula $\Phi \in \mathcal{G}^+_\wedge(\mathcal{T}, 0, s, Arith_c)$ there exists a formula $\Phi' \in \mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$ such that $mod_u(\Phi) = mod_u(\Phi')$.*

Proof. We first prove that some arithmetical operators and constants, including those of $Arith_c$ (for a given predefined constant c), are definable in $\mathcal{G}^+_\wedge(0, s, c)$. The precise statement of Lemma 3 (definability of these objects in $\mathcal{G}^+_\wedge(0, s)$ when c is itself definable in $\mathcal{G}^+_\wedge(0, s)$) will trivially follows.

To proceed, we give for each arithmetical object $\mathcal{O}_i$ $(1 \leq i \leq 12)$ to be expressed in $\mathcal{G}^+_\wedge$:

- its *definition*, that is the way it must be interpreted on each finite domain $n = \{0, \ldots, n-1\}$;

– when the definition can not be directly translated into a formula of $\mathcal{G}^+_\wedge$, we give an alternative definition, that is a property that fully *characterises* the object under consideration ;
– Then we give a formula $def_i(\mathcal{O}_i)$ in $\mathcal{G}^+_\wedge(0, s, \mathcal{O}_i)$ which defines $\mathcal{O}_i$, *i.e.* such that $\langle n, 0, s, \mathcal{O}_i\rangle \models def_i$ if, and only if, $\mathcal{O}_i$ meets the definition above mentionned.

Actually, this last assertion is not totally correct, since the formula defining a given object $\mathcal{O}_i$ will sometimes involve some objects $\mathcal{O}_j$, $j < i$, previously defined. That means that we rather have to consider the series $(def_i)_{1\leq i\leq 12}$ as a whole, the conjunction of all these formulas defining both the object $\mathcal{O}_1, \dots,$ $\mathcal{O}_{12}$.

Until the end of this proof, we denote by $+, ., -$ the usual corresponding operation on $\mathbb{N}$.

Recall that our successor function is not cyclic, *i.e.* it is interpreted on each finite domain n by: $\forall x < n-1$, $s(x) = x+1$ and $s(n-1) = n-1$.

1. Constants **max** et **max − 1** :
Definition. $\max = n-1$ and $\max -1 = n-2$.
Characterization. $\max -1 \neq \max$ and $s(\max -1) = s(\max) = \max$.
Definability.

$$def_1 \equiv \exists \mathcal{X} \ : \ \mathcal{X}(\max) = 1 \wedge \mathcal{X}(\max -1) = 0 \wedge s(\max -1) = s(\max) = \max .$$

2. Function **p** :
Definition. $\forall x < n$, $p(x) = \begin{cases} x-1 & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$
Characterization. $\forall x \neq 0$, $sp(x) = x$ and $p(0) = 0$ and $p(n-1) = n-2$.
Definability.

$$def_2 \equiv \forall x : sps(x) = s(x) \wedge p(0) = 0 \wedge p(\max) = \max -1.$$

3. Function $+_c$:
Definition. $s^c(x) = \begin{cases} x+c & \text{if } x+c < n \\ n-1 & \text{else} \end{cases}$
Characterization. s^c is the only function that commute with s and that maps 0 onto c.
Definability.

$$def_3 \equiv \forall x : s^c s(x) = ss^c(x) \wedge s^c(0) = c$$

4. Constant **max − c** :
Definition. $c = (n-1) - c$
Characterization. $\max -c$ is the smallest x such that $s^c(x) = \max$.
Definability.

$$def_4 \equiv s^c(\max -c) = \max \wedge s^c p(\max -c) = \max -1.$$

5. Function $\mathbf{p^c}$:
Definition. $p^c(x) = \begin{cases} x-c & \text{if } x-c \geq 0 \\ 0 & \text{else} \end{cases}$
Characterization. p^c is the only function that commute with p and that maps max onto $\max -c$.
Definability.

$$def_5 \equiv \forall x : p^c p(x) = pp^c(x) \wedge p^c(\max) = \max -c.$$

6. Function $\boldsymbol{\pi_c}$:
Definition. $\forall x < n,\ \pi_c(x) = \begin{cases} c-x & \text{if } x \leq c \\ 0 & \text{else} \end{cases}$
Characterization. $\pi_c(0) = c$ and $\forall x > 0,\ \pi_c(x+1) = \pi_c(x) - 1$.
Definability.

$$def_6 \equiv \forall x : \pi_c s(x) = p\pi_c(x) \wedge \pi_c(0) = c.$$

7. Function $\boldsymbol{\mathcal{X}_c}$, for $c \neq \max$
Definition. $\forall x < n,\ \mathcal{X}_c(x) = 0$ if $x \neq c$ and $\mathcal{X}_c(c) = 1$.
Characterization. $\forall x < c : \mathcal{X}_c(x) = 0$ and $\forall x > c : \mathcal{X}_c(x) = 0$ and $\mathcal{X}(c) = 1$. We also use the equalities: $[0, c[= \{\pi_c s(x),\ x < n\}$ and $]c, n-1] = \{s^c s(x),\ x < n\}$.
Definability.

$$def_7 \equiv \forall x : \mathcal{X}_c \pi_c s(x) = \mathcal{X}_c s^c s(x) = 0 \wedge \mathcal{X}_c(c) = 1.$$

8. Functions $\boldsymbol{\mathcal{X}_{\max}}$:
Definition. $\forall x < n,\ \mathcal{X}_{\max}(x) = 0$ if $x \neq \max$ and $\mathcal{X}_{\max}(\max) = 1$.
Definability.

$$def_8 \equiv \forall x : \mathcal{X}_{\max} p(x) = 0 \wedge \mathcal{X}_{\max}(\max) = 1.$$

9. Functions $\boldsymbol{\mathcal{X}_{<c}}$ for $c \neq 0$:
Definition. $\forall x < n,\ \mathcal{X}_{<c}(x) = \begin{cases} 1 & \text{if } x < c \\ 0 & \text{else} \end{cases}$
Definability.

$$def_9 \equiv \forall x : \mathcal{X}_{<c} \pi_c s(x) = 1 \wedge \mathcal{X}_{<c} s^c(x) = 0.$$

10. Function $\times_c$:
Definition. $\forall x < n,\ M_c(x) = \begin{cases} cx & \text{if } cx < n \\ n-1 & \text{else} \end{cases}$
Definability.

$$def_{10} \equiv \forall x : M_c s(x) = s^c M_c(x) \wedge M_c(0) = 0.$$

11. Function $\mathcal{Q}_c$:
Definition. $\forall x < n,\ Q_c(x) = \lfloor x/c \rfloor$.
Characterization. $\forall x < c,\ Q_c(x) = 0$ and $\forall x \leq (n-1) - c,\ Q_c(x+c) = Q_c(x) + 1$.
Definability.

$$def_{11} \equiv \forall x : Q_c \pi_c s(x) = 0 \wedge Q_c s^c \pi_{\max -c}(x) = s Q_c \pi_{\max -c}(x).$$

12. Function $\mathcal{R}_c$:
Definition. $\forall x < n$, $R_c(x) = x - cQ_c(x)$.
Characterization. $\forall x < c$, $R_c(x) = x$ and $\forall x \leq (n-1)-c$, $R_c(x+c) = R_c(x)$.
Definability.

$$def_{12} \equiv \forall x \,:\, R_c\pi_c s(x) = x \wedge R_c s^c \pi_{\max - c}(x) = R_c \pi_{\max - c}(x).$$

Now, assume that c is definable in $\mathcal{G}^+_\wedge(0, s)$, *i.e.* that there exists a formula $\phi \in \mathcal{G}^+_\wedge(a, 0, s)$ (where a is a constant symbol) such that for any finite domain n and any $a \in n$, $n \models \phi(a, o, s)$ if, and only if, a is the interpretation of c in n.

Then for any unary signature $\mathcal{T}$, each formula $\Phi \in \mathcal{G}^+_\wedge(\mathcal{T}, 0, s, Arith_c)$ is equivalent on unary structures to the formula:

$$\Phi' \equiv \exists c, \max, \max -1, p, \ldots, Q_c, R_c \;:\; \phi(c) \wedge \bigwedge_{1 \leq i \leq 12} def_i \wedge \Phi,$$

which belongs to $\mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$. QED

Remark. 1. The previous lemma still holds when replacing $\mathcal{G}^+_\wedge$ by $\mathcal{G}$.
2. From now on, we will denote by the usual notation the operators s^c and M_c (*i.e.* $a + c$ and ca instead of $s^c(a)$ and $M_c(a)$) and we will no longer deal with arithmetic operators on $\mathbb{N}$.
3. In the previous proof, we have dealt with constant symbols, although we had claimed before that we will only use unary function symbols. Actually, this is only for convenience, since the occurences of the "constant" 0 can be replaced everywhere by the term $0(x)$, and since any formula $\exists c\phi \in \mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$ (where c is a constant symbol) can be translated into $\exists \tilde{c}\tilde{\phi}$, where $\tilde{c}$ is a unary function symbol, and $\tilde{\phi}$ is obtained from ϕ by the substitution $c \rightsquigarrow \tilde{c}(0(x))$.
4. Let c be a predefined constant symbol. Then the formula $\exists \boldsymbol{g}\, (\forall x < c)\, \phi(x)$ (resp. $\exists \boldsymbol{g}\, (\forall x \geq c)\, \phi(x)$) is equivalent, over unary structures, to the formula $\exists \boldsymbol{g}\, \forall x\, \phi(c - s(x))$ (resp. $\exists \boldsymbol{g}\, \forall x\, \phi(c + x)$), which can, in turn, be written in $\mathcal{G}^+_\wedge(0, s, c)$.

3 Suppression of the linear order and the negation

We prove here that, given any unary signature $\mathcal{T}$, the class $\mathcal{G}(\mathcal{T}, 0, s, <)$ reduces to the class $\mathcal{G}^+(\mathcal{T}, 0, s)$. The supression of negation (*i.e.* the reduction of $\mathcal{G}(\mathcal{T}, 0, s, <)$ to $\mathcal{G}^+(\mathcal{T}, 0, s, <)$) is trivial. So the core of our proof will consist in defining the linear order with unary functions without reintroducing the negation.

Lemma 4. *Let $\mathcal{T}$ be a unary signature. For each $\Psi \in \mathcal{G}(\mathcal{T}, 0, s, <)$, there exists a formula $\Psi^+ \in \mathcal{G}^+(\mathcal{T}, 0, s)$ such that $mod_u(\Psi) = mod_u(\Psi^+)$.*

Proof. Let us consider a formula $\Psi \in \mathcal{G}(\mathcal{T}, 0, s, <)$. We can assume that Ψ is a positive formula (*i.e.* belongs to $\mathcal{G}^+(\mathcal{T}, 0, s, <)$) since each negative subformula $\neg(\sigma(x) = \tau(x))$ (resp. $\neg(\sigma(x) < \tau(x))$) of Ψ can be replaced by its positive translation $\sigma(x) < \tau(x) \vee \tau(x) < \sigma(x)$ (resp. $\tau(x) < \sigma(x) \vee \sigma(x) = \tau(x)$).

Let b, π_0, π_1, π_2 be the predefined symbols whose interpretation on every domain $n \in \mathbb{N}$ is determinated by the assertions:

- $b = \lfloor \sqrt{n-1} \rfloor$;
- $\forall x < n : \; x = \pi_0(x) + \pi_1(x)b + \pi_2(x)b^2$;
- $\forall x < n, \; \forall i < 3 : \; \pi_i(x) < b$.

If $n > 2$, then $n - 1 < b^3$ and the π_i are defined on every $x < n$. Furthermore, these symbols are characterized by the following facts:

- $b^2 \leq n - 1 < (b+1)^2$, which can be also written: $b^2 - 1 < n - 1 \leq b^2 + 2b$;
- for each $x < n$, $\pi_2(x) = Q_{b^2}(x)$, $\pi_0(x) + \pi_1(x)b = R_{b^2}(x)$, $\pi_1(x) = Q_b(\pi_0(x) + \pi_1(x)b)$ and $\pi_0(x) = R_b(\pi_0(x) + \pi_1(x)b)$.

Consequently, they are definable in $\mathcal{G}^+(0,s)$ (and even in $\mathcal{G}^+_\wedge(0,s)$) by the formulas:

- $p(b^2) \neq \max \wedge b(b+2) = \max$ (recall that each operator of our arithmetic maps on max every $x < n$ whose image by the associated "natural" operator (*i.e.* defined on $\mathbb{N}$) should have been larger than $n - 1$) ;
- $\forall x : \; \pi_2(x) = \mathcal{Q}_{b^2}(x) \; \wedge \; \pi_1(x) = \mathcal{Q}_b \mathcal{R}_{b^2}(x) \; \wedge \; \pi_0(x) = \mathcal{R}_b \mathcal{R}_{b^2}(x)$.

Thus, for any unary signature $\mathcal{T}$, each formula in $\mathcal{G}^+(\mathcal{T}, 0, s, b, \boldsymbol{\pi})$ can be considered as belonging to $\mathcal{G}^+(\mathcal{T}, 0, s)$.

Now, let us consider the unary function symbol Inf, interpreted on each domain $n \in \mathbb{N}$ by the function:

$$\forall x < n : \; \mathrm{Inf}(x) = \begin{cases} 1 & \text{if } \; \pi_0(x) < \pi_1(x); \\ 0 & \text{else.} \end{cases}$$

To convince ourselves that this function is definable in $\mathcal{G}^+(0,s)$, let us consider the following picture, which represents (by colored dots) the set $\mathcal{I} = \{x < n : \; \mathrm{Inf}(x) = 1\}$:

$$\begin{array}{lllll} 0 & b & 2b & b^2 & n-1 \\ \circ\circ\ldots\circ\circ & \bullet\circ\ldots\circ\circ & \bullet\bullet\ldots\circ\circ\ldots\bullet\bullet\ldots\bullet & \circ\circ & \ldots\ldots\;\circ \end{array}$$

We can clearly characterize this set with the assertions:

- $\mathcal{I}$ contains b;
- for each $x < (b-1)b$ belonging to $\mathcal{I}$, $x + b$ and $x + b + 1$ are in $\mathcal{I}$;
- $\mathcal{I}$ contains exactly $\frac{(b-1)b}{2}$ elements.

This leads to the definition of Inf by the following formula:

$$\mathrm{Inf}(b) = 1 \; \wedge$$

$$(\forall x < (b-1)b)(\mathrm{Inf}(x) = 0 \vee \mathrm{Inf}(x+b) = \mathrm{Inf}(x+b+1) = 1) \; \wedge$$

$$\exists f, f', c \; \forall x : \; ff'(x) = x \; \wedge \; c = \tfrac{(b-1)b}{2} \; \wedge \; \mathrm{Inf}f(x) = \mathcal{X}_{<c}(x)$$

This definability of Inf in $\mathcal{G}^+(0,s)$ allows to simulate the restriction to $[0,b-1]$ of the linear order $<$. More precisely, if two integers $x, y < n$ are smaller than b, the assertion $x < y$ can be translated into:

$$\exists t: \ \pi_0(t) = x \wedge \pi_1(t) = y \wedge \mathrm{Inf}(t) = 1.$$

We will denote by $x <_b y$ this formula. Consequently, the formula $\sigma(x) < \tau(x)$, where $\sigma, \tau \in \{\boldsymbol{f}, \boldsymbol{g}, 0, s\}^*$, is equivalent to:

$$\begin{cases} \pi_2(\sigma(t)) <_b \pi_2(\tau(t)) \\ \vee \\ \pi_2(\sigma(t)) = \pi_2(\tau(t)) \ \wedge \ \begin{cases} \pi_1(\sigma(t)) <_b \pi_1(\tau(t)) \\ \vee \\ \pi_1(\sigma(t)) = \pi_1(\tau(t)) \ \wedge \ \pi_0(\sigma(t)) <_b \pi_0(\tau(t)) \end{cases} \end{cases}$$

which claims that the tuple $(\pi_2\sigma(x), \pi_1\sigma(x), \pi_0\sigma(x))$ is strictly smaller than $(\pi_2\tau(x), \pi_1\tau(x), \pi_0\tau(x))$ according to the lexicographical order over $b \times b \times b$. We wil denote by $\boldsymbol{\pi}(\sigma(x)) \prec \boldsymbol{\pi}(\tau(x))$ this formula. Remark that $\boldsymbol{\pi}(\sigma(x)) \prec \boldsymbol{\pi}(\tau(x))$ has a prenex form of type $\exists \boldsymbol{t}\phi$, where $\boldsymbol{t}$ is a tuple of first-order variable and ϕ is quantifier free and positive.

Let us now come back to our formula $\Psi \equiv \exists \boldsymbol{g} \forall x \psi$ in $\mathcal{G}^+(\boldsymbol{f}, 0, s, <)$. Because of the form of the formulas $\boldsymbol{\pi}(\sigma(x)) \prec \boldsymbol{\pi}(\tau(x))$, the formula $\exists \boldsymbol{g} \forall x \tilde{\psi}$, where $\tilde{\psi}$ is obtained from ψ by the substitutions $\sigma(x) < \tau(x) \rightsquigarrow \boldsymbol{\pi}(\sigma(x)) \prec \boldsymbol{\pi}(\tau(x))$, belongs to $\mathcal{G}^+(0,s)$ and is equivalent to Ψ. It is the sought formula Ψ^+.

4 Definition by cases

The goal of this section is to prove the definability in $\mathcal{G}^+_\wedge(0,s)$ of the function q defined on domain n by:

$$\forall x < n, q(x) = \begin{cases} q_0(x) & \text{if} \ \ r(x) = 0; \\ \vdots \\ q_{k-1}(x) & \text{if} \ \ r(x) = k-1, \end{cases}$$

where r and the q_i's are given unary functions, r mapping n into $[0, k-1]$. the difficulty is limited to the case $k = 2$, *i.e.* to the case of a function q defined by:

$$\forall x < n, q(x) = \begin{cases} q_0(x) & \text{if} \ \ r(x) = 0; \\ q_1(x) & \text{if} \ \ r(x) = 1. \end{cases} \tag{1}$$

The other cases $(k > 2)$ easily follow by recurrence on k.

So let us first assume that $k=2$ and consider four functions q, q_0, q_1, r meeting (1) and such that $\forall x < n$, $r(x) \in \{0,1\}$.

Let us furthermore assume that $r^{-1}\langle 0\rangle \neq \emptyset$ and $r^{-1}\langle 1\rangle \neq \emptyset$, where $r^{-1}\langle 0\rangle$ stands for $\{x < n : r(x) = 0\}$ (and similarly for $r^{-1}\langle 1\rangle$), and set $c = \mathrm{card}(r^{-1}\langle 0\rangle)$.

To define q conjunctively, we would like to write:

$$\forall x \in r^{-1}\langle 0\rangle : \; q(x) = q_0(x) \quad \text{and} \quad \forall x \in r^{-1}\langle 1\rangle : \; q(x) = q_1(x).$$

But of course, quantifications like $(\forall x \in r^{-1}\langle 0\rangle)$ are not allowed in our syntax. Nevertheless, we can simulate them thanks to a bijection $b : n \to n$ that precisely maps $[0, c[$ onto $r^{-1}\langle 0\rangle$ and $[c, \max]$ onto $r^{-1}\langle 1\rangle$. Indeed, such a bijection beeing defined, any assertion $\forall x \in r^{-1}\langle 0\rangle : \; \phi(x)$ (resp. $\forall x \in r^{-1}\langle 1\rangle : \; \phi(x)$), where ϕ is a conjunction of equalities, can be translated into $\forall x < c : \; \phi(b(x))$ (resp. $\forall x \geq c : \; \phi(b(x))$), and we know that such formulas can be written in $\mathcal{G}^{+}_{\wedge}(0, s)$.

Finally, we easily deduce, for the case where $\boldsymbol{r^{-1}\langle 0\rangle \neq \emptyset}$ **and** $\boldsymbol{r^{-1}\langle 1\rangle \neq \emptyset}$:

$$\forall x < n, q(x) = \begin{cases} q_0(x) & \text{if } \; r(x) = 0; \\ q_1(x) & \text{if } \; r(x) = 1. \end{cases}$$
$$\text{iff}$$
$$\langle n, q, q_0, q_1, r, 0, s\rangle \models \left(\begin{array}{c} \exists b, b' \exists c : \\ \forall x : bb'(x) = x \;\wedge \\ \forall x < c : rb(x) = 0 \wedge qb(x) = q_0 b(x) \;\wedge \\ \forall x \geq c : rb(x) = 1 \wedge qb(x) = q_1 b(x). \end{array} \right)$$

The case $\boldsymbol{r^{-1}\langle 0\rangle = \emptyset}$ is very simpler. We just have to write:

$$\forall x < n, q(x) = \begin{cases} q_0(x) & \text{if } \; r(x) = 0; \\ q_1(x) & \text{if } \; r(x) = 1. \end{cases}$$
$$\text{iff}$$
$$\langle n, q, q_0, q_1, r, 0, s\rangle \models \forall x : r(x) = 1 \; \wedge q(x) = q_1(x).$$

And for the case where $\boldsymbol{r^{-1}\langle 1\rangle = \emptyset}$:

$$\forall x < n, q(x) = \begin{cases} q_0(x) & \text{if } \; r(x) = 0; \\ q_1(x) & \text{if } \; r(x) = 1. \end{cases}$$
$$\text{iff}$$
$$\langle n, q, q_0, q_1, r, 0, s\rangle \models \forall x : r(x) = 0 \; \wedge q(x) = q_0(x).$$

To describe this three cases with a single conjunctive formula, we build a formula $\Theta(\varepsilon_0, \varepsilon_1, \varepsilon_2)$, where the ε_i's are three constants behaving as *swiches* which branch Θ on the good interpretation according to the behavior of r.

More precisely, consider the three formulas:

$$\Theta_0 \equiv \forall x : \varepsilon_0 \cdot r(x) = \varepsilon_0 \cdot 0 \wedge \varepsilon_0 \cdot q(x) = \varepsilon_0 \cdot q_0(x)$$
$$\Theta_1 \equiv \forall x : \varepsilon_1 \cdot r(x) = \varepsilon_1 \cdot 1 \wedge \varepsilon_1 \cdot q(x) = \varepsilon_1 \cdot q_1(x)$$
$$\Theta_2 \equiv \exists b, b' \exists c : \left(\begin{array}{c} \forall x : bb'(x) = x \;\wedge \\ \forall x < c : \varepsilon_2 \cdot rb(x) = \varepsilon_2 \cdot 0 \wedge \varepsilon_2 \cdot qb(x) = \varepsilon_2 \cdot q_0 b(x) \;\wedge \\ \forall x \geq c : \varepsilon_2 \cdot rb(x) = \varepsilon_2 \cdot 1 \wedge \varepsilon_2 \cdot qb(x) = \varepsilon_2 \cdot q_1 b(x). \end{array} \right)$$

Then

$$\text{CASE}_2 \equiv \exists \varepsilon_0, \varepsilon_1, \varepsilon_2 : \ \varepsilon_0 + \varepsilon_1 + \varepsilon_2 = 1 \ \wedge \ \Theta_0(\varepsilon_0) \wedge \Theta_1(\varepsilon_1) \wedge \Theta_2(\varepsilon_2).$$

Lemma 5. *For each integer $k \geq 2$, there exists a formula CASE_k in $\mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$, where $\mathcal{T}$ is a $(k+2)$-tuple of unary function symbols, such that: for every $n \geq k$ and every $(k+2)$-tuple $(q, q_0, \ldots, q_{k-1}, r)$ of unary functions on n satisfying: $\forall x < n$, $r(x) < k$, we have:*

$$n \models \text{CASE}_k(q, q_0, \ldots, q_{k-1}, r, 0, s)$$
$$\textit{iff}$$
$$\forall x < n, q(x) = \begin{cases} q_0(x) & \textit{if} \ \ r(x) = 0; \\ \vdots & \\ q_{k-1}(x) & \textit{if} \ \ r(x) = k-1. \end{cases}$$

Actually, we will write this formula $q = \text{CASE}_k(q_0, \ldots, q_{k-1}, r, 0, s)$, instead of $\text{CASE}_k(q, q_0, \ldots, q_{k-1}, r, 0, s)$.

Proof. We prove this result by recurrence over k. The fact that the recurrence hypothesis holds for $k = 2$ has just been proved. It just remains to prove the recurrence step in building CASE_{k+1} from CASE_2, $\ldots$, CASE_k.

Let $q, q_0, \ldots, q_k : n \to n$ and $r : n \to [0 \ldots k]$ be unary functions on a domain $n \geq k$ such that:

$$\forall x < n, \ q(x) = \begin{cases} q_0(x) & \text{if} \ \ r(x) = 0; \\ \vdots & \\ q_{k-1}(x) & \text{if} \ \ r(x) = k-1; \\ q_k(x) & \text{if} \ \ r(x) = k \end{cases}$$

Let us now consider the functions $r', q' : n \to n$ defined by:

$$\forall x < n, \ r'(x) = \begin{cases} r(x) & \text{if} \ \ \mathcal{X}_k r(x) = 0; \\ 0 & \text{if} \ \ \mathcal{X}_k r(x) = 1 \end{cases}$$

and

$$\forall x < n, \ q'(x) = \begin{cases} q_0(x) & \text{if} \ \ r'(x) = 0; \\ \vdots & \\ q_{k-1}(x) & \text{if} \ \ r'(x) = k-1. \end{cases}$$

Since the functions $\mathcal{X}_k r$ and r' respectively take their values in $\{0, 1\}$ and $\{0, \ldots, k-1\}$, the functions r' and q' are definable in $\mathcal{G}^+_\wedge(0, s)$ by the formulas: $r' = \text{CASE}_2(r, 0 \ ; \ \mathcal{X}_k r)$ and $q' = \text{CASE}_k(q_0, \ldots, q_{k-1} \ ; \ r')$.

Moreover, q is definable from q and r' by:

$$\forall x < n, \ q(x) = \begin{cases} q'(x) & \text{if} \ \ \mathcal{X}_k r(x) = 0; \\ q_k(x) & \text{if} \ \ \mathcal{X}_k r(x) = 1 \end{cases}$$

Consequently, the formula

$$\begin{aligned}\exists r', q' \ : \ & r' = \text{CASE}_2(r, 0 \ ; \ \mathcal{X}_k r) \ \wedge \\ & q' = \text{CASE}_k(q_0, \ldots, q_{k-1} \ ; \ r') \ \wedge \\ & q = \text{CASE}_2(q', q_k \ ; \ \mathcal{X}_k r)\end{aligned}$$

defines q and is in $\mathcal{G}^+_\wedge(0, s)$. It is the sought formula $q = \text{CASE}_{k+1}(q_0, \ldots, q_k \ ; \ r)$.

From now on we will omit the suffix k in the notation $q = \text{CASE}_k(q_0, \ldots, q_{k-1} \ ; \ r)$, the context (and specifically the arity of the tuple $(q_0, \ldots, q_{k-1})$) allowing to make it explicit.

5 Main result

Theorem 6. *Let $\mathcal{T}$ be a unary signature. For every $\Phi \in \mathcal{G}(\mathcal{T}, 0, s)$ there exists $\Phi_\wedge \in \mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$ such that $mod_u(\Phi) = mod_u(\Phi_\wedge)$.*

Proof. Let us first assume that $\boldsymbol{h}$ is any tuple of unary function symbols and that $(\sigma_i)_{i<k}$ and $(\tau_i)_{i<k}$ are two k-tuple of words over the alphabet $\boldsymbol{g}$. Then for any domain $n \in \mathbb{N}^*$, the following assertions are equivalent:

(i) $\langle n, \boldsymbol{h}, 0, s \rangle \models \forall x : \ \sigma_0(x) = \tau_0(x) \vee \ldots \vee \sigma_{k-1}(x) = \tau_{k-1}(x)$;

(ii) $\langle n, \boldsymbol{h}, 0, s \rangle \models \exists r, \tilde{\sigma}, \tilde{\tau} \ :$

$$\begin{aligned} & \forall x : \ r(x) < k \\ & \wedge \ \forall x : \ \tilde{\sigma}(x) = \begin{cases} \sigma_0(x) & \text{if} \quad r(x) = 0; \\ \quad \vdots & \\ \sigma_{k-1}(x) & \text{if} \quad r(x) = k-1. \end{cases} \\ & \wedge \ \forall x : \ \tilde{\tau}(x) = \begin{cases} \tau_0(x) & \text{if} \quad r(x) = 0; \\ \quad \vdots & \\ \tau_{k-1}(x) & \text{if} \quad r(x) = k-1. \end{cases} \\ & \wedge \ \forall x : \ \tilde{\sigma}(x) = \tilde{\tau}(x). \end{aligned}$$

The implication $(ii) \Rightarrow (i)$ is clear. For the implication $(i) \Rightarrow (ii)$, take the function r defined by: $\forall x$,

$$r(x) = \begin{cases} 0 \ \text{ if } \ \sigma_0(x) = \tau_0(x); \\ 1 \ \text{ if } \ \sigma_0(x) \neq \tau_0(x) \wedge \sigma_1(x) = \tau_1(x); \\ 2 \ \text{ if } \ \sigma_0(x) \neq \tau_0(x) \wedge \sigma_1(x) \neq \tau_1(x) \wedge \sigma_2(x) = \tau_2(x); \\ \vdots \\ k-1 \ \text{ if } \ \sigma_0(x) \neq \tau_0(x) \wedge \ldots \wedge \sigma_{k-2}(x) \neq \tau_{k-2}(x) \wedge \sigma_{k-1}(x) = \tau_{k-1}(x) \end{cases}$$

Remark that, thanks to the previous section, the formula quoted in (ii) can be formally written in $\mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$ as:

$$\exists r, \tilde{\sigma}, \tilde{\tau} : \ \forall x (r(x) < k) \ \wedge$$

$$\tilde{\sigma} = \text{CASE}(\sigma_0, \ldots, \sigma_{k-1}; r) \ \wedge \ \tilde{\tau} = \text{CASE}(\tau_0, \ldots, \tau_{k-1}; r) \ \wedge$$

$$\forall x (\tilde{\sigma}(x) = \tilde{\tau}(x)).$$

We will denote by $\Theta(\boldsymbol{\sigma}, \boldsymbol{\tau})$ this formula.

Now, consider a formula $\Phi \equiv \exists \boldsymbol{g} \forall x \bigwedge_{j<\ell} \bigvee_{i<k} \sigma_{ji}(x) = \tau_{ji}(x)$ of $\mathcal{G}^+(\mathcal{T}, 0, s)$. Rewrite Φ as $\exists \boldsymbol{g} \bigwedge_{j<\ell} \forall x \bigvee_{i<k} \sigma_{ji}(x) = \tau_{ji}(x)$ and replace each conjunct $\forall x \bigvee_{i<k} \sigma_{ji}(x) = \tau_{ji}(x)$ by the equivalent formula $\Theta(\boldsymbol{\sigma}_j, \boldsymbol{\tau}_j)$ (where $\boldsymbol{\sigma}_j$ stands for $(\sigma_{j0}, \ldots, \sigma_{j(k-1)})$ for each $j < \ell$, and similarly for $\boldsymbol{\tau}_j$). The obtained formula $\exists \boldsymbol{g} \bigwedge_{j<\ell} \Theta(\boldsymbol{\sigma}_j, \boldsymbol{\tau}_j)$ has a prenex form in $\mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$: it is the sought formula $\Phi_\wedge$.

We are now going to state the conjunctive normalization of the Grandjean's characterization of NLIN. We won't detail the precise definition of NLIN. (It can be found in [1].) Just recall that it is the class of set of unary structures that can be recognized in linear time by a nondeterministic RAM (under uniform cost measure).

Theorem 7 (Grandjean, [1]). *A set $\mathcal{S}$ of $\mathcal{T}$-unary structures is in NLIN if and only if there exists a formula $\Phi \in \mathcal{G}(\mathcal{T}, 0, s, <)$ such that $\mathcal{S} = mod_u(\Phi)$.*

Corollary 8. *A set $\mathcal{S}$ of $\mathcal{T}$-unary structures is in NLIN if and only if there exists a formula $\Phi \in \mathcal{G}^+_\wedge(\mathcal{T}, 0, s)$ such that $\mathcal{S} = mod_u(\Phi)$.*

Conclusion

Our main result can be understood as a "strong skolemization" of formulas of the form $\exists \boldsymbol{g} \forall x \exists \boldsymbol{y}\, \phi$, where $\boldsymbol{g}$ is a tuple of unary functions, $\boldsymbol{y}$ is a tuple of first-order variables, and ϕ is a quantifier-free formula.

For such formulas indeed, we can supress all kind of disjunction: those related to the existential first-order quantification, thanks to the usual skolemization, and those appearing in the matrix ϕ, tanks to our reduction.

The interest of Corollary 8 is strengthen by the belonging to NLIN of most of the natural NP-complete problems (notably the 21 problems listed by Karp in [2]). Thus, all these problems are characterized by a formula of $\mathcal{G}^+_\wedge(0, s)$. Concerning the other problems of NP, a generalization of Corollary 8 state the following (see proposition 2.17 in [3]):

> *A set $\mathcal{S}$ of unary structures is in $NTIME_{RAM}(n^k)$ if and only if there exists a formula Φ of the form $\exists \boldsymbol{G} \forall \boldsymbol{x} \phi$, where $\boldsymbol{G}$ is a tuple of k-ary functions, $\boldsymbol{x}$ is a k-tuple of first-order variables and ϕ is a quantifier-free conjunction of equalities, such that $\mathcal{S} = mod_u(\Phi)$.*

Acknoledgements

I wish to express my thanks to Etienne Grandjean for many stimulating conversations and to Thomas Schwentick for suggesting me a determining simplification of the proof of Theorem 6.

References

1. E. Grandjean, F.Olive, *Monadic Logical Definability of NP-Complete Problems*, CSL'94 Proc. LNCS 933 (1995), pp. 190-204.
2. R.M. Karp, *Reducibility among combinatorial problems*, IBM Symp. 1972, Complexity of Computers Computations, Plenum Press, New York, 1972.
3. F. Olive, *Caractérisation Logique des problèmes NP : Robustesse et Normalisation*, Thèse de Doctorat, 1996, Université de Caen, France.