Rewriting,

between computer science and algebra

Pierre-Louis Curien

(CNRS – Paris 7 – INRIA)

2/4/2011, Tsinghua University, Beijing

(improved) 9/5/2011, South China Normal University, Guangzhou

(few corrections made 19/2/2012)

Rewriting and theory of computation (1/2)

The λ-calculus : **Church** (around 1930). A remarkably simple syntax :

$$M ::= x \mid M M \mid \lambda x.M$$

and just one *rewriting rule*, β-reduction :

$$\boxed{(\lambda x.M)N \longrightarrow M[x \leftarrow N]}$$

This reduction can be applied inside a term :

$$\frac{M \longrightarrow M'}{\lambda x.M \longrightarrow \lambda x.M'} \quad \frac{M \longrightarrow M'}{M N \longrightarrow M' N} \quad \frac{N \longrightarrow N'}{M N \longrightarrow M N'}$$

or, equivalently,

$$C[(\lambda x.M)N] \longrightarrow C[M[x \leftarrow N]]$$

where C is a context = term with a hole. The subterm $(\lambda x.M)N$ is called a redex.

Rewriting and theory of computation (2/2)

λ-calculus was the **first formalism** for the notion of computable function (before Turing !) :

– Natural numbers encoded by *Church numerals*

$$\underline{n} = \lambda f.(\lambda x.f(\ldots(f(x)\ldots)) \quad (f \text{ applied } n \text{ times})$$

– $F : \mathbb{N} \times \cdots \times \mathbb{N} \to \mathbb{N}$ is computable if there exists M such that for all $n_1, \ldots, n_k$:

$$(\ldots(M\underline{n_1})\ldots\underline{n_k}) \longrightarrow^* \underline{F(n_1, \ldots, n_k)}$$

where $\longrightarrow^*$ is the reflexive-transitive closure of $\longrightarrow$.

This supposes that β-reduction is deterministic : if $(M\underline{n_1})\ldots\underline{n_k} \longrightarrow^* \underline{p}$ and $(M\underline{n_1})\ldots\underline{n_k} \longrightarrow^* \underline{q}$, then $\underline{p} = \underline{q}$.

Church-Rosser property

Church-Rosser (1935) : β-reduction is **confluent** :

$$\text{If } M \longrightarrow^* N_1 \text{ and } M \longrightarrow^* N_2, \text{ then } \exists\, N \; N_1 \longrightarrow^* N \text{ and } N_2 \longrightarrow^* N$$

Equivalent formulation

$$\text{If } N_1 \longleftrightarrow^* N_2, \text{ then } \exists\, N \; N_1 \longrightarrow^* N \text{ and } N_2 \longrightarrow^* N$$

where $\longleftrightarrow^*$ is the reflexive-symmetric-transitive closure of $\longrightarrow$.

The equivalence is proved by an easy diagram chasing at the level of abstract rewriting systems, i.e., oriented graphs.

The second formulation is often called "the Church-Rosser property".

Rewriting in computer science (1/2)

Another key property is termination : no infinite reduction sequences. Then every term M has a (non necessarily unique) normal form : $M \longrightarrow M_1 \longrightarrow^* M_n$, where M_n contains no redex.

Termination of β holds only for typed lambda-calculi.

(Effective) termination + confluence = convergence lead to decidability of equality. To check $N_1 = N_2$, compute normal forms P_1 for N_1 and P_2 for N_2 and check whether P_1 and P_2 coincide.

Weak termination = the existence of a normal form for every term is thus enough.

Rewriting in computer science (2/2)

Newman (1942) (abstract) **diamond property** : local confluence and termination imply confluence

Knuth-Bendix (1970) (term rewriting)

– One can restrict the verification of local confluence to the (finitely many) critical pairs (= minimal redex overlappings)

– One can complete the rewriting system, adding new rewrite rules without changing the equational theory

The case of β-reduction is easy (except that the theory has to be lifted to higher-order rewriting) : no critical pairs.

An example of Knuth-Bendix completion : the group axioms (1/2)

We start with

$$(R1)\ (x*y)*z \to x*(y*z) \quad (R2)\ i(x)*x \to 1 \quad (R3)\ 1*x \to x$$

We add successively ($(Ri) - (Rj)$ identifies the relevant critical pair) :

- $(R4)\ i(x)*(x*y) \to y$ (from $(R1) - (R2)$)
- $(R1')\ i(1)*x \to x$ (from $(R1) - (R4)$)
- $(R2')\ i(i(x))*1 \to x$ (from $(R2) - (R4)$)
- $(R3')\ i(i(x))*y \to x*y$ (from $(R1) - (R2')$)
- $(R5)\ x*1 \to x$ (from $(R2') - (R3')$)
- $(R6)\ i(i(x)) \to x$ (from $(R2') - (R5)$) (remove $(R2'), (R3')$, now redundant)
- $(R7)\ i(1) \to 1$ (from $(R1') - (R5)$) (remove $(R1')$)
- $(R8)\ x*i(x) \to 1$ (from $(R2) - (R6)$)
- $(R9)\ x*(i(x)*y) \to y$ (from $(R1) - (R8)$)
- $(R4')\ x*(y*i(x*y)) \to 1$ (from $(R1) - (R8)$)
- $(R5')\ y*i(x*y) \to i(x)$ (from $(R4) - (R4')$)
- $(R10)\ i(x*y) \to i(y)*i(x)$ (from $(R4) - (R5')$) (remove $(R4'), (R5')$)

An example of Knuth-Bendix completion : the group axioms (2/2)

The system of the ten rules $(R1), \ldots, (R10)$ is convergent (clever choice of orientation each time a rule is added !).

Consequence. We can present the free group explicitly rather than merely as a quotient. The normal forms provide representatives of the equivalence classes of terms.

The normal forms are in one-to-one correspondence with the words u made from the generators a, b, c and their formal inverses a^{-1}, b^{-1} such that no subword of the form aa^{-1} or $a^{-1}a$ occur in u (which is the classical presentation of the free group).

Proof : analyse the shapes of normal forms !

An aside : Homology of rewriting

Squier (1987). A word rewriting system $\mathcal{R}$ on an alphabet Σ induces a complex of abelian groups

$$\mathbb{Z}[\mathcal{T}] \xrightarrow{\partial_4} \mathbb{Z}[\mathcal{P}] \xrightarrow{\partial_3} \mathbb{Z}[\mathcal{R}] \xrightarrow{\partial_2} \mathbb{Z}[\Sigma] \xrightarrow{\partial_1} \mathbb{Z}$$

where $\mathcal{P}$ is the set of critical pairs, $\mathcal{T}$ is the set of critical triples (minimal situations where a redex overlaps with two other redexes), and where

$$\partial_1 = 0 \quad , \quad \partial_2 = \text{sourge} - \text{target} \quad , \quad \partial_3 = \text{left path} - \text{right path}$$

– The homology of this complex depends only on the quotient monoïd (i.e. does not depend on the presentation).
– If $\mathcal{R}$ is convergent, then the third homology group of this complex is finitely generated.

This theorem provides an invariant allowing to show that certain (even decidable) monoids do not have a convergent presentation.

There are extensions of this theory to term rewriting, and to higher-dimensional rewriting (Malbos, Guiraud).

Rewriting in algebra

– **Janet** (1920) (systems of linear pde's) *Janet bases*
– **Shirshov** (1962) (Lie algebras, etc...) (Bokut, Chen Yuqun, Chen Yongshan,...)
– **Hironaka** (1964) *standard bases*
– **Buchberger** (1965) *Gröbner bases*
– **Bergman** (1978) (establishes the link with Knuth-Bendix and Newman)

These seem to be largely independent works.

And recently :

– **Dotsenko-Khoroshkin** (2010) *Gröbner bases for operads*

Gröbner bases (1/3)

We follow the presentation of the following nice book by F. Baader and T. Nipkow : Term rewriting and all that, Cambridge Univ. Press (1998)

We work in $\mathbb{K}[X_1, \ldots, X_n]$ (polynomials). Suppose given a total order on monomials (which should be a congruence for multiplication and should contain the division relation).

Let $R = \{f_1, \ldots, f_k\} \subseteq \mathbb{K}[X_1, \ldots, X_n]$. We are interested in the following decision problem : given f, is f in $\langle R \rangle$ (the ideal generated by R) ?

We can write (up to dividing by a scalar) each f_i as $f_i = m_i - r_i$, where all monomials of r_i are $< m_i$, et see R as a rewriting system $\mathcal{R}$ on polynomials :

$$R = \{f_1, \ldots, f_k\} \qquad \mathcal{R} = \{m_1 \to r_1 \ , \ \ldots m_k \to r_k\}$$

Gröbner bases (2/3)

$$\boxed{\dfrac{f = am' \oplus g \qquad m' = m''m \qquad m \to r \in \mathcal{R}}{f \to am''r + g}}$$

(where $\oplus$ emphasizes the polynomial f as formal sum of monomials while $+$ denotes an addition of polynomials).

If the rule is $m_i \to r_i$, this rephrases as

$$f \to f - am''f_i \quad (= \text{ a step in the division of } f \text{ by } f_i)$$

The rewriting relation terminates (one alway replaces a monomial by a collection of strictly lower monomials). A normal form of f can be read as the remainder of **a** division :

$$f = (h_1 f_1 + \ldots + h_k f_k) + r$$

where no $< m_i$ divides a monomial of $r = f - (h_1 f_1 + \ldots + h_k f_k)$.

Gröbner bases (3/3)

If all critical pairs are confluent, $\{f_1, \ldots, f_k\}$ is called a *Gröbner basis*.

Other terminologies : all ambiguities are resolved (Bergmann), all results of compositions reduce to 0 (Shirshov), all S-polynomials reduce to 0 (Buchberger).

If R is a Gröbner base, then the above membership problem is decidable.

Si R is not a Gröbner basis, it can be completed : Buchberger (in a way which predates Knuth-Bendix).

Buchberger's algorithm

One executes the following loop : Look for critical pairs. When the corresponding S-polynomial does not reduce to 0, add its normal form to the set of rules. This in turn creates new critical pairs, etc....

Termination : Because only a normal form with respect to the current set R_n of rules is added, its leading monomial is not a multiple of any other left-hand side of R_n, and in particular is not a multiple of the leading monomials of previously introduced new rules.

If the algorithm goes on for ever, it must add new rules for ever (as checking all critical pairs takes a finite number of steps). But then the associated sequence of their leading monomials contradicts the well-partial-order structure of $Mon[X_1, \ldots, X_n]$ (more on this later).

Note that we do not see a quantity decreasing. The termination argument is indirect in that sense. With Janet bases, we shall see a more explicit form of termination. (The same holds for the refined version of Buchberger's algorithm where one not only adds rules, but removes rules that become redundant.)

An example of Buchberger completion

$$(R1)\ X_1^2 X_2 \to X_1^2 \qquad (R2)\ X_1 X_2^2 \to X_2^2$$

We add :

– $(R3)\ X_1^2 \to X_2^2$ (from $(R1) - (R2)$)

– $(R4)\ X_2^3 \to X_2^2$ (from $(R1) - (R3)$)

This gives the following set of normal forms : $1, X_1, X_2, X_1X_2, X_2^2$.

Warning : X_1, X_2, the indeterminates, play a different role from variables in term rewriting systems : here the rules are more like ground rewriting rules (rules involving no variables). We use capitalised letters to stress the difference.

Gröbner bases vs Poincaré-Birkhoff-Witt bases

If $R = \{f_1, \ldots, f_k\}$ is a Gröbner basis, then the vector space V spanned by the monomials in normal form is $\cong \mathbb{K}[X_1, \ldots, X_n]/\langle R \rangle$. This basis of normal forms is called Poincaré-Birkhoff-Witt basis (PBW basis for short).

Proof :

1. Every polynomial g writes as $g = \mathrm{nf}(g) + (g - \mathrm{nf}(g))$, hence

$$\mathbb{K}[X_1, \ldots, X_n]/\langle R \rangle = V + \langle R \rangle$$

Lemma (holds for any R) : If $f - g \in \langle R \rangle$, then $f \leftrightarrow^* g$. This follows from

– $f_i \to 0$ (by definition of $\to$)

– $\leftrightarrow^*$ is a congruence for the addition and multiplication of polynomials (a consequence of the following property : if $f \to g$, then for any k there exists l such that $f + k \to^{\leq 1} l$ and $g + k \to^{\leq 1} l$ ($^{\leq 1}$ means “at most one step”)).

2. The sum is direct : if $g_1, g_1' \in V$, $g_2, g_2' \in \langle R \rangle$ and $g_1 + g_2 = g_1' + g_2'$, then $g_1 - g_1' \in \langle R \rangle$, hence by the lemma $g_1 = \mathrm{nf}(g_1) = \mathrm{nf}(g_1') = g_1'$.

Another example (non commutative polynomials)

Let $\mathcal{L}$ be a Lie algebra, with basis $\{v_1, v_2, \ldots\}$, i.e. a vector space on this basis, endowed with a Lie bracket, satisfying :

anti-symmetry	$[x, y] = -[y, x]$
Jacobi	$[[x, y], z] + [[y, z], x] + [[z, x], y] = 0$

The universal enveloping algebra over $\mathcal{L}$ is

$$U(\mathcal{L}) = T(\mathcal{L})/\langle R \rangle \qquad \text{where } R = \{v_j v_i - v_i v_j + [v_i, v_j] \mid i < j\}$$

Poincaré-Birkhoff-Witt :

$$U(\mathcal{L}) \quad \cong \quad S(\mathcal{L})$$

as vector spaces where $S(\mathcal{L}) = T(\mathcal{L})/\langle \{v_j v_i - v_i v_j \mid i < j\} \rangle$ is the symmetric algebra.

We set $v_i v_j < v_j v_i$, for all $i < j$.

R is a Gröbner base

$$
\begin{tikzcd}[column sep=tiny]
 & v_k v_j v_i \quad (i<j<k) \arrow[dl] \arrow[dr] & \\
v_j v_k v_i - [v_j, v_k] v_i \arrow[d] & & v_k v_i v_j - v_k [v_i, v_j] \arrow[d] \\
v_j v_i v_k - v_j [v_i, v_k] - [v_j, v_k] v_i \arrow[d] & & v_i v_k v_j - [v_i, v_k] v_j - v_k [v_i, v_j] \arrow[d] \\
v_i v_j v_k - [v_i, v_j] v_k - v_j [v_i, v_k] - [v_j, v_k] v_i \arrow[dr] & & v_i v_j v_k - v_i [v_j, v_k] - [v_i, v_k] v_j - v_k [v_i, v_j] \arrow[dl] \\
 & \text{(using Jacobi and antisymmetry)} &
\end{tikzcd}
$$

Completing the proof of PBW theorem

The normal forms are the monomials where no $v_i v_j$ $(i < j)$ occurs : But this is the basis for $S(\mathcal{L})$!

That we can prove the PBW theorem through Gröbner bases justifies the terminology of PBW bases.

Gröbner bases in mathematical textbooks (1/3)

We follow Ufnarowskij (Combinatorial and asymptotic methods in algebra, in Algebra VI, Encycopedia of Mathematical Sciences, Springer, 1995), but see also Eisenbud (Commutative algebra with a view toward algebraic geometry, Springer, 1994)

The definition (and equivalent to the one above) of Gröbner base there
– does depend on the choice of an ordering of monomials,
– but is not "algorithmic" (no reduction, no division).

Let m, n be two monomials. Let $d_m(n)$ be the number of occurrences of m in n. This extends to polynomials, taking their leading monomials. This also extends to $d_F(g)$ (take the sum of the $d_f(g)$, $f \in F$).

Gröbner bases in mathematical textbooks (2/3)

Suppose an order has been fixed. Let $R = \{f_1, \ldots, f_n\}$ (with associated rewriting system $\mathcal{R}$). The following are equivalent :

1. All critical pairs of $\mathcal{R}$ are confluent
2. $\mathcal{R}$ is confluent (or, equivalently, convergent, or Church-Rosser, or is such that every term has a unique normal form)
3. $K[X_1, \ldots, X_n] = \langle R \rangle \oplus V$, where V is spanned by the monomials in normal form wrt $\mathcal{R}$
4. for all non null element f of $\langle R \rangle$, $d_R(f) > 0$

- $(1) \Rightarrow (2)$ is Newman's lemma (cf. slide 5).
- We have proved $(2) \Rightarrow (3)$. For the converse, suppose that f has two normal forms f_1, f_2. Then $f = f_1 + (f - f_1) = f_2 + (f - f_2)$, and hence $f_1 = f_2$.
- $(2) \Rightarrow (4)$. Since $f - 0 \in \langle R \rangle$, we have $f \leftrightarrow^* 0$, hence $f \rightarrow^* 0$, and the leading monomial cannot be left untouched (cf. standardisation arguments in rewriting theory)
- $(4) \Rightarrow (3)$. This relies on the following two properties :
  - under condition (4) monomials in normal form coincide with normal monomials, defined as follows : m is normal if $d_{\langle R \rangle} m = 0$.
  - For any R, we have $K[X_1, \ldots, X_n] = \langle R \rangle \oplus W$, where W is spanned by the normal monomials of $K[X_1, \ldots, X_n]$.

Gröbner bases in mathematical textbooks (3/3)

We prove $K[X_1, \ldots, X_n] = \langle R \rangle \oplus W$ as follows :

- The sum is direct. Suppose that $f \in W \cap \langle R \rangle$. Then a fortiori its leading monomial is normal. But the leading monomial of a polynomial in $\langle R \rangle$ cannot be normal by definition of a normal monomial.
- One proves by induction on the order on monomials that every monomial can be decomposed. Let m be a monomial. There are two cases :

  1. m is normal. Then $m = 0 + m$ provides a decomposition.
  2. m is not normal. Thus there exists $f \in \langle R \rangle$ with m' dividing m as leading monomial. Then (because the order is assumed to be a congruence !), by multiplying with $\frac{m}{m'}$ we may assume that both $m' = m$ and m is the leading monomial of f. Therefore f writes as $f = \alpha m + g$, where g's monomials are all $< m$. Hence we can apply induction to g (monomial-wise) and we get $f = \alpha m + g + h$, where $f, g \in \langle R \rangle$ and $h \in W$. So $m = \alpha^{-1}(f - g) - \alpha^{-1} h$ does the job.

Byproduct : a fifth equivalent definition of Gröbner bases

5. For any $f \in \langle R \rangle$, there exists **a** reduction path $f \to^* 0$.

- $(1) \Rightarrow (5)$: Remember from the lemma on slide 15 that $f \in \langle R \rangle$ iff $f \leftrightarrow^* 0$. Then by confluence $f \in \langle R \rangle$ iff $f \to^* 0$ (for all reduction paths, and hence a fortiori for some reduction path).
- $(5) \Rightarrow (4)$: Let $f \in \langle R \rangle \setminus \{0\}$, and $f \to^* 0$ through some reduction path. Then, again, the leading polynomial of f cannot be left untouched by the reduction. Hence $f \to^* g \to h \to^* 0$, where $d_R(f) = d_R(g)$ and the redex reduced in the step $g \to h$ is the leading monomial of g, hence $d_R(g) > 0$. Wrapping up, we have $d_R(f) > 0$.

Two flavours of "confluence"

Buchberger's algorithm actually checks a variation of condition 1. So we have actually a 6th equivalent condition :

1'. If $h \to f$, $h \to g$ form a critical pair, then there exists a reduction path $f - g \to^* 0$.

We sketch the proof that (1') is equivalent to (1). That (1') is implied follows directly from condition (5). That (1') implies (1) follows from the following lemma :

Lemma : If $f - g \to h$, then there exist f_1, g_1 such that

$$f \to^{\leq 1} f_1 \ , \ g \to^{\leq 1} g_1 \ , \text{ and } f_1 - g_1 = h$$

Let $f = amm_i \oplus f' \qquad g = bmm_i \oplus g' \qquad h = (a-b)mr_i + f' - g'$. Then we have $f \to^{\leq 1} amr_i + f'$ (0 steps if $a = 0$) and $f \to^{\leq 1} bmr_i + g'$, and we conclude.

Topics for discussion

Common generalisations

Transfer of techniques from algebra to rewriting and conversely

Further historical quest (**Janet**, Hironaka,...)