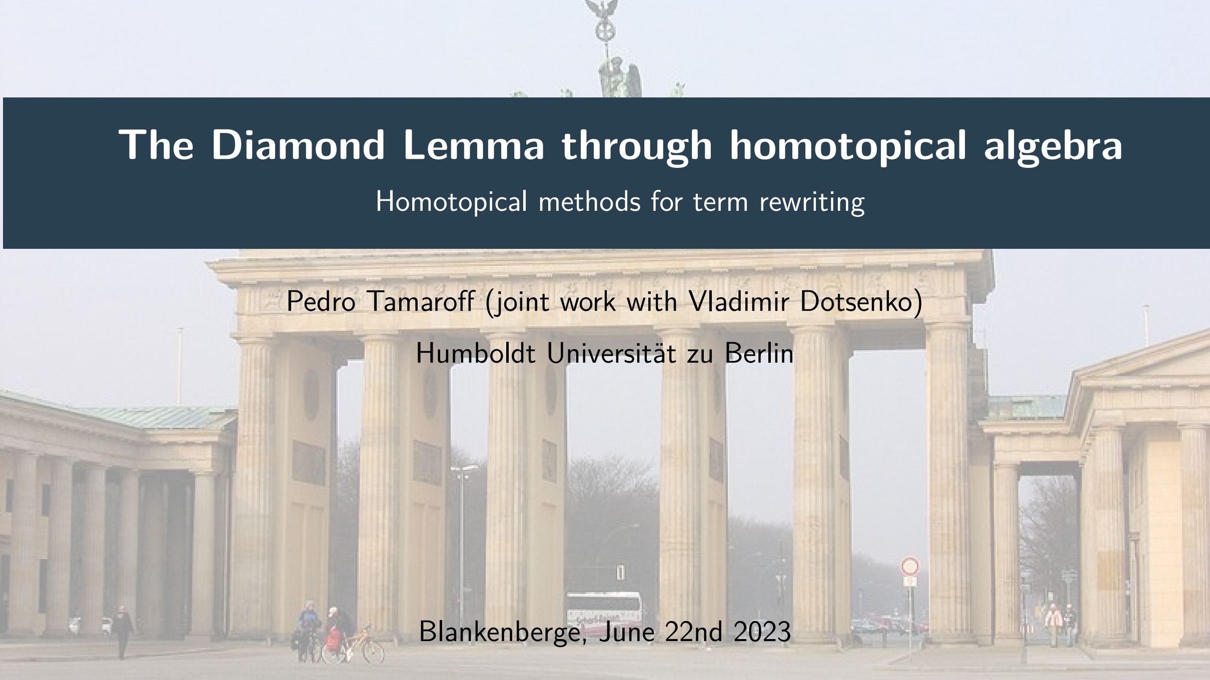

The Diamond Lemma through homotopical algebra

Homotopical methods for term rewriting

Pedro Tamaroff (joint work with Vladimir Dotsenko)

Humboldt Universität zu Berlin

Blankenberge, June 22nd 2023

What's our goal?

Term rewriting theory: offer 'homotopical' intuition about the Diamond Lemma and improvements, such as the Triangle Lemma; obtain a guiding principle to generalize them for richer algebraic structures.

Homotopical algebra: offer an abstract explanation to an effective (but perhaps technical) test that guarantees uniqueness of normal forms using "overlapping ambiguities".

Resolutions and invariants through term rewriting

One can obtain information from an algebra A by looking at its Yoneda coalgebra $\mathrm{Tor}^A(\mathbb{k},\mathbb{k})$. *How do we compute it?*

1. (Green–Happel–Zacharia '85) Compute projective resolutions of the simple modules of monomial algebras, later the (co)algebra structure.
2. (Anick '86, Bardzell '97, Chouhy–Solotar '14) Compute projective resolutions of (bi)modules over algebras with a Gröbner basis or a confluent rewriting system.
3. (Sköldberg '05, T. '19, Dotsenko-Gélinas-T. '19) Compute Hochschild cohomology of an algebra using its Yoneda algebra, and other richer invariants.

Leading terms and overlappings

Intuitively, an algebra determined by generators and relations can be ‘pinned down’ by a good choice of leading terms. *How do we formalize this?*

1. (Ufnarovsky ’82) A procedure for determining the growth of an associative algebra through a graph constructed from a Gröbner basis.
2. (Chuang–King 05’, T. ’18) Resolutions of monomial algebras in terms of overlappings; (T ’18) Conjectural description in the case of Gröbner bases.
3. (Green–Hille–Schroll ’20) Affine variety associated to each rewriting system on a free algebra. Generic points are the monomial algebras.

Rewriting systems

A rewriting system is a triple $\Sigma = (X, W, f)$ where $W \subseteq \langle X \rangle$ is a set of words and $f : W \longrightarrow T(X)$ is a 'rewriting function'.

We like to think of f as a collection of *rewriting rules* $w \rightsquigarrow f(w)$, or as a collection or *relations* $\Gamma_f = \{w - f(w) : w \in W\} \subseteq T(X)$.

Rewriting systems

A rewriting system is a triple $\Sigma = (X, W, f)$ where $W \subseteq \langle X \rangle$ is a set of words and $f : W \longrightarrow T(X)$ is a 'rewriting function'.

We like to think of f as a collection of *rewriting rules* $w \rightsquigarrow f(w)$, or as a collection or *relations* $\Gamma_f = \{w - f(w) : w \in W\} \subseteq T(X)$.

Each rewriting Σ system has its own *associated algebra* $A_\Sigma = T(X)/(\Gamma_f)$. A central problem to begin studying this algebra is:

Basis problem

How can we single out a $\Bbbk$-linear basis of monomials of A_Σ using the datum of Σ?

Normal forms

A *basic reduction* is a triple $\rho = (u, w, v)$ where $w \in W$ and $u, v \in \langle X \rangle$. It defines $\rho : T(X) \longrightarrow T(X)$ that fixes every basis monomial except that $\rho(uwv) = uf(w)v$.

Careful! We could have $uwv = u'wv'$ even though $u' \neq u$, so a basic reduction assumes a *choice of occurrence of* w.

Normal forms

A *basic reduction* is a triple $\rho = (u, w, v)$ where $w \in W$ and $u, v \in \langle X \rangle$. It defines $\rho : T(X) \longrightarrow T(X)$ that fixes every basis monomial except that $\rho(uwv) = uf(w)v$.

Careful! We could have $uwv = u'wv'$ even though $u' \neq u$, so a basic reduction assumes a *choice of occurrence of* w.

Definition

A monomial is irreducible if it is fixed by every basic reduction. A normal form is a linear combination of irreducible monomials.

Goal: a basis of normal forms. *Problem:* the projection from normal forms to A_Σ may not be surjective nor injective.

Termination

Given a monomial that is not irreducible, we may rewrite it. If at some point we reach a normal form, we are done.

Definition

The system Σ is terminating if every infinite chain of reductions stabilizes on each basis monomial.

Equivalently, if and only if the projection is *surjective*. We can guarantee termination using compatible partial orders.

Compatible orders

A partial order on $\langle X \rangle$ is:

- *Multiplicative*: if each left and right multiplication is increasing.
- *Compatible* with Σ: if $w > w'$ whenever w' appears in $f(w)$.
- *Noetherian*: if it is a well-order (desc. chain condition holds).

Compatible orders

A partial order on $\langle X \rangle$ is:

- *Multiplicative*: if each left and right multiplication is increasing.
- *Compatible* with Σ: if $w > w'$ whenever w' appears in $f(w)$.
- *Noetherian*: if it is a well-order (desc. chain condition holds).

Fact

The system Σ is terminating if and only if it admits a multiplicative compatible Noetherian partial order.

Example: if $X = \{x, y, z\}$ and $z^2 \rightsquigarrow xy$, this system is terminating, as rewriting decreases the number of appearances of z.

Confluence

Let us assume that $T(X)^{\mathrm{irr}} \longrightarrow A_\Sigma$ is onto. How can injectivity fail?

Example: if $X = \{x, y, z\}$ and $z^2 \rightsquigarrow xy$, then $zxy - xyz = (z^2 - xy)z - z(z^2 - xy)$, so the map is not injective[1].

Solution: add the rule $zyx \rightsquigarrow xyz$. This fixes the previous step. But what if it adds new 'ambiguities'?

[1] A better choice is $xy \rightsquigarrow z^2$.

Confluence

Let us assume that $T(X)^{\mathrm{irr}} \longrightarrow A_\Sigma$ is onto. How can injectivity fail?

Example: if $X = \{x, y, z\}$ and $z^2 \rightsquigarrow xy$, then $zxy - xyz = (z^2 - xy)z - z(z^2 - xy)$, so the map is not injective[1].

Solution: add the rule $zyx \rightsquigarrow xyz$. This fixes the previous step. But what if it adds new 'ambiguities'?

Definition

We say Σ is confluent if the projection is injective, and that it is convergent if it is both confluent and terminating.

[1] A better choice is $xy \rightsquigarrow z^2$.

Discovering the Diamond Lemma

Problem: what is an effectively verifiable criterion to determine confluence?

Let us consider the rewriting system $\Sigma_0 = (X, W, 0)$, where $0 : W \longrightarrow T(X)$ sends every word to the zero element.

In this case, convergence is trivial: we have a $\Bbbk$-linear isomorphism $A_{\Sigma_0} \longrightarrow T(X)^{\mathrm{irr}}$, and multiplication of two normal monomials is either zero or normal.

Note!

The system Σ is convergent if and only if the map $A_{\Sigma_0} \longrightarrow A_\Sigma$ sending a normal monomial to its coset is a vector space isomorphism.

Homotopy theory of associative algebras

We observed that Σ is convergent precisely when A_Σ is a deformation of A_{Σ_0}: the underlying vector spaces are isomorphic, and only the multiplication table is modified.

Question: is there any way of deducing convergence *solely* in terms of A_{Σ_0} from this deformation theoretic point of view?

The object that controls the deformation theory of an associative algebra is obtained by first 'resolving' the algebra itself.

Idea: replace equalities by a witness of this equality: our relations $w - f(w) = 0$ will be replaced by a witness z such that $dz = w - f(w)$.

Multiplicative resolutions

The end result of continuing this process builds a *resolution* of our algebra, which one usually calls a model.

Definition

A model of A is a quasi-isomorphism $(TV, d) \longrightarrow A$ where (TV, d) is free and d is triangulated. We call (TV, d) a *Sullivan algebra*.

That d be triangulated means V decomposes into summands V_k for which $dV_k = T(V_{<k})$, which means (TV, d) is built by attaching cells in an ordered way. *Good for inductive arguments!*

An anchor to build resolutions

A good starting point to build resolutions is a nc variant of the Koszul complex.

Definition

(Shafarevich '64, Golod '88) The Shafarevich complex $Ш(A, S)$ of the datum $(A, S \subseteq A)$ is the free product $(T(U) * A, d)$ where $U = \Bbbk S[1]$ and $d(e_s) = s \in A$.

If we take $A = T(X)$ and $S = \Gamma_0$ where $\Sigma = (X, W, 0)$ then we will write the complex by $Ш_{\Sigma_0}$. By construction this is a Sullivan algebra with $H_0(Ш_{\Sigma_0}) = A_{\Sigma_0}$, but in general there are higher non-trivial cycles, which we need to kill.

Resolutions adapted to systems

Definition

We say a resolution (F, d) of A_{Σ_0} is an acyclic extension of $Ш_{\Sigma_0}$ if it coincides with it in degrees zero and one and it is graded by $\langle X \rangle$.

The first main result is that a resolution of the trivial rewriting Σ system can be perturbed into a resolution of Σ if and only if the latter if convergent.

Recall that a perturbation of (TV, d) is a DGA algebra of the form $(TV, d + d')$. We say d' is Σ-compatible if it lowers the $\langle X \rangle$-grading with respect to our chosen order $<_\Sigma$.

Theorem I

Theorem (Dotsenko–T.)

The system Σ is convergent if and only if any acyclic extension of $Ш_{\Sigma_0}$ admits a Σ-compatible perturbation which is an acyclic extension of $Ш_{\Sigma}$.

Spelling this out, this means that we can find some d' such that $d'(v)$ is a linear combination of elements smaller than $v \in V$, and that $dd' + d'd + d'^2 = 0$.

Moreover, it means that $d'|V_1 : \Bbbk W \longrightarrow T(X)$ must coincide with the unique linear extension of $-f : W \longrightarrow T(X)$.

Can we make this effective?

Although perturbation theory is very elegant, and produces intricate combinatorial equations to produce deformations, constructing acyclic extensions by hand is markedly difficult.

Moreover, the usual assumption that is needed to use its formulas is precisely the condition we want to check.

However, we have at hand an obstruction theory to the existence of perturbations —without having to compute them— written in the language of dg Lie algebras.

Perturbations are MC elements

We will state and prove a 'model dependent' criterion for convergence.

Definition

A perturbation of a model (TV, d) is the datum of a degree -1 derivation $F : TV \longrightarrow TV$ such that $(d+F)^2 = 0$.

Expanding, we see that F defines a perturbation if and only if

$$\nabla F = \partial F + \frac{1}{2}[F, F] = 0$$

where $\partial = \mathrm{Ad}_d$.

This is the ubiquitous *Maurer–Cartan equation* in the Lie algebra of derivations of the DGA algebra (TV, d).

When this is an acyclic extension of Ш_{Σ_0}, let us denote the resulting Lie algebra of derivations by $(\mathfrak{g}(\Sigma_0), \partial)$, so that $F \in \mathrm{MC}(\mathfrak{g}(\Sigma_0))$.

The conditions of our Theorem I imply, in particular, that:

1. $F|V_1 : \Bbbk W \longrightarrow T(X)$ coincides with $-\Bbbk f$.
2. $F|V_2$ is Σ-compatible.
3. F satisfies the MC equation in degree two.

Theorem II

Our next theorem implies that this is enough to obtain an acyclic extension of $Ш_\Sigma$ from one of $Ш_{\Sigma_0}$.

Theorem (Dotsenko-T.)

The rewriting system Σ is convergent if and only if there exists $F \in \mathfrak{g}(\Sigma_0)$ with the properties above.

The proof proceeds by induction, constructing an acyclic extension of $Ш_\Sigma$ by modifying F into a bona-fide Maurer–Cartan element.

The Diamond Lemma revisited

Let us now spell out what these three conditions mean, and how they are effectively verifiable. As a first step, we can rediscover the notion of an ‘overlapping ambiguity’.

For (TV, d) an acyclic extension of Ш_{Σ_0}, let us pick some $c \in V_2$. The element dc is of degree one, and thus belongs to $TV_0 \otimes V_1 \otimes TV_0$:

$$dc = \sum_{w \in W} u \otimes w \otimes v.$$

Definition

We define the *obstruction* S_c as the element $S_c = \sum_{w \in W} u \cdot f(w) \cdot v \in TV_0$.

The Diamond Lemma revisited

Our third result is that these are indeed the model dependent obstructions to convergence, and hence give a ‘coordinate independent’ version of the Diamond Lemma of Bergman.

Theorem (Dotsenko-T.)

The rewriting system Σ is convergent if and only if for each word homogeneous element $c \in V_2$, the obstruction S_c can be rewritten to zero using a sequence of reductions from Σ.

The usual Diamond Lemma

As before, fix X and W. Dotsenko and Khoroshkin ('13) build a resolution of $T\langle X|W\rangle$ by analysing the occurrence of divisors in a monomial.

By looking at the shape of elements c in degree two, one can compute that the obstructions S_c defined above are *exactly* S-polynomials corresponding to inclusion and overlap ambiguities, which recovers Bergman's Diamond Lemma on the nose.

The Triangle Lemma

Using the notion of a ‘chain’ (appearing in work of Anick, and before this in work of Green, Happel and Zacharia and implicitly in the work of Backelin) we constructed a resolution of $T\langle X|W\rangle$ (T. ’18).

By looking at the shape of elements c in degree two, one can compute that the obstructions S_c defined above are exactly S-polynomials corresponding to *essential* overlap ambiguities; this recovers the optimization of the Diamond Lemma known as the Triangle Lemma.

Other algebraic structures

The inclusion exclusion model of Dotsenko–Khoroshkin works equally well in the case of algebraic operads.

The proof of the Diamond Lemma we obtained carries through, mutatis mutandis, to recover the appropriate version for non-symmetric and shuffle operads.

Thank you!