THE DIAMOND LEMMA FOR RING THEORY

NOTES BY APOORVA KHARE
PRESENTED TO AKAKI TIKARADZE IN ECKHART 16 ON AUG 16, 2007, THURSDAY
AND TO DMITRIY BOYARCHENKO IN RYERSON ON NOV 26, 2008, WEDNESDAY

It is useful to keep an example in mind, while discussing the entire theory. So we use the following two examples; we work in both cases over a commutative ground ring k.

(1) Let $\mathfrak{sl}_2$ be the three-dimensional semisimple split Lie algebra over k. Then we claim that $\{f^a h^b e^c : a, b, c \geq 0\}$ form a k-basis of $\mathfrak{U}(\mathfrak{sl}_2)$.
(2) More generally, let $\mathfrak{g}$ be any Lie algebra that is k-free, with some fixed basis $\{x_i : i \in I\}$. Choose and fix some total order $\prec$ on I (by Zorn's Lemma). Then

$$\mathcal{I} := \{x_{i_1}^{j_1} \dots x_{i_k}^{j_k} : k, j_1, \dots, j_k \geq 0,\ i_1 \prec i_2 \prec \dots \prec i_k\}$$

is a k-basis of $\mathfrak{U}\mathfrak{g}$. In other words, we will prove the usual Poincare-Birkhoff-Witt Theorem for Lie algebras (of any dimension, over any field) as a consequence of the Diamond Lemma.

0.1. **Introduction.** In general, suppose X is a set of *letters* or *generators*, and $\langle X \rangle$ the monoid (including the empty word 1) generated by X. That is, if V is the k-span of X, then $k\langle X \rangle = T_k(V)$ is the free associative k-algebra generated by X.

We have *relations* $\{W_\sigma - f_\sigma : \sigma \in \mathcal{S}\}$, where

$$W_\sigma \in \langle X \rangle,\ f_\sigma \in k\langle X \rangle\ \forall \sigma \in \mathcal{S}$$

and we want to determine a k-form (or k-basis) for the ring or k-algebra $R = k\langle X \rangle / I$, where I is the two-sided ideal generated by $\{W_\sigma - f_\sigma : \sigma \in \mathcal{S}\}$.

For example, in the example above, $X = \{f, h, e\}$ and one can take

$$\mathcal{S} := \{ef - (fe + h),\ eh - (he - 2e),\ hf - (fh - 2f)\}$$

or in the general case, $X = \{x_i : i \in I\}$, and one can take

$$\mathcal{S} := \{x_i x_j - (x_j x_i + a_{ij}) : i \succ j \in I\}$$

where $a_{ij} := [x_i, x_j] \in \mathfrak{g}$.

0.2. **Reductions.** Crucial to this whole process is the concept of a *reduction*. This is simply the following: given $\sigma \in \mathcal{S}$ and $A, B \in \langle X \rangle$, the *$AW_\sigma B$-reduction* is the k-module endomorphism of $k\langle X \rangle$ that sends $AW_\sigma B \mapsto Af_\sigma B$, and every other element of $\langle X \rangle$ to itself (and is extended by k-linearity).

Note that given a general element in $\langle X \rangle$ or in $k\langle X \rangle$, there exist more than one reductions that can be applied to it.

Finally, a word is *irreducible* if every reduction leaves it unchanged. Let $\langle X \rangle_{\text{irr}}$ denote the set of irreducible words. So in our example, $\langle X \rangle_{\text{irr}} = \{f^a h^b e^c : a, b, c \geq 0\}$, or in the general case, $\langle X \rangle_{\text{irr}} = \mathcal{I}$ (which was defined above).

0.3. **The Diamond Lemma in Graph Theory.** The original Diamond Lemma was stated in Graph Theory: given a directed graph, if it satisfies the

- *DCC (Descending Chain Condition)* - namely, that any directed path from a point has finite length;
  and the
- *Diamond Condition* - namely, given any point in the graph, any two distinct directed edges starting at this point, may be extended to directed paths that end up at the same point,

then every connected graph component has a unique "maximal" vertex.

0.4. **Setup for Bergman's analogue of the Diamond Lemma.** What Bergman does in [Be] is to prove an analogous theorem for rings in the above setup - but he also proves it with weaker assumptions. Here are the analogies to the Diamond Lemma in Graph Theory:

- In place of the directed graph, he requires a *partial order* on the set $\langle X \rangle$, that is compatible with: $B \leq B' \Rightarrow ABC \leq AB'C$ for all $B, B', A, C \in \langle X \rangle$. This is called a *semigroup partial order* on $\langle X \rangle$. Perhaps this can be thought of as saying that the graph structure (locally) is "translation-invariant".

  For instance, here's a partial order on $X = \{f, h, e\} : f < h < e$, and words of smaller length are smaller. In the general example, the order on X comes from that on I (so we still call it $\prec$), and words of smaller length are smaller. Both these (for the former is a special case of the latter) easily extend to a semigroup partial order, namely, the *lexicographic order* on X.

- In place of a (directed) *edge*, he now desires to put the above notion of reductions. Thus, any substring of any monomial that equals W_σ, amounts to one possible reduction for that monomial.

  A *path* is simply an ordered collection of edges, that is now analogous to a sequence of reductions. In order that the path (in this setup) be *directed*, Bergman requires that the partial order is *compatible with the reductions*. In other words, every monomial in f_σ should be strictly less than the monomial $W_\sigma \in \langle X \rangle$, for each $\sigma \in \mathcal{S}$.

In our examples, it is easily verified that the given partial order is compatible with $\mathcal{S}$.

- "Maximal" vertices, of course, are ones that have no directed edge starting at them - so the analogous notion is precisely that of irreducible words. In other words, Bergman wants to show that
  (1) every expression in $k\langle X \rangle$ reduces to a unique irreducible one; and
  (2) the set of irreducible words (under the DCC and Diamond Conditions) precisely gives the desired k-form for $R = k\langle X \rangle / I$.

0.5. **Descending Chain Condition.** The analogy involves the notion of *reduction-finiteness*; namely, that every monomial in $\langle X \rangle$, when subjected to any sequence of reductions, eventually stabilizes; in other words, it is irreducible beyond some finite number of steps (reductions).

Typically, this is proved via the notion of a *misordering index*, denoted by $mis : \langle X \rangle \to \mathbb{Z}_{\geq 0}$, such that every reduction strictly reduces its value. Namely, $mis(w) < mis(W_\sigma)$ for each monomial w in f_σ (for all $\sigma \in \mathcal{S}$).

For example, in our examples, given $w = x_1 \dots x_n$ with $x_i \in X \ \forall i$, define the pseudo-misorder to be

$$mis'(x_1 \dots x_n) := \#\{(i,j) : i < j,\ x_i \succ x_j\}$$

where we recall that the partial order $\prec$ on X is the same as that in I, via the bijection $i \mapsto x_i$. Now define the actual misorder to be

$$mis(x_1 \dots x_n) := mis'(x_1 \dots x_n) + n^2$$

Let us show that the descending chain condition holds. For some reduction to apply on such a word w as above, we assume that for some fixed i, we have $x_i \succ x_{i+1}$. Then $[x_i, x_{i+1}]$ is one of the a_{rs}'s for some $r, s \in I$, so we write it as a k-linear sum of monomials: $[x_i, x_{i+1}] = \sum_{j=1}^{l} a_j x_j$, where $x_j \in X, a_j \in k$ for all j.

Also write $w_1 = x_1 \dots x_{i-1}$ and $w_2 = x_{i+2} \dots x_n$, for convenience. Then the reduction involving x_i and x_{i+1} says

$$w_1 x_i x_{i+1} w_2 \mapsto w_1 x_{i+1} x_i w_2 + \sum_{j=1}^{l} a_j (w_1 x_j w_2)$$

It is easy to see that $mis(w) > mis(w_1 x_{i+1} x_i w_2)$, so one only needs to show that each of the summands in the second expression, has a strictly smaller misordering index than w. But if we let $w' = w_1 w_2$, then $mis'(w_1 x_j w_2) \leq mis'(w') + (n-2)$, because the only "extra" pairs in the pseudo-misorder of $w_1 x_j w_2$, must involve x_j. Moreover, we clearly have $mis'(w') < mis'(w)$, given that $x_i \succ x_{i+1}$.

Putting all this together, along with the definition of mis in terms of mis', we get

$$\begin{aligned} mis(w_1 x_j w_2) &= mis'(w_1 x_j w_2) + (n-1)^2 \\ &\leq mis'(w') + (n-2) + (n-1)^2 < mis'(w) + n^2 - n - 1 \\ &< mis(w) \end{aligned}$$

Hence we have shown that the misordering index strictly reduces with each reduction in this case.

0.6. **Diamond Condition, and minimal ambiguities.** This means that given any expression $w \in k\langle X\rangle$, if more than one reduction is (nontrivially) applicable, then any infinite sequence of reductions yields the same answer (and note that if DCC holds, then all these sequences stabilise beyond some finite point; if a misordering index can be constructed, then there is a uniform bound on the point of stabilizing as well).

This is where Bergman strengthens the Diamond Lemma in graph theory, in two different ways:

(1) One only needs to resolve ambiguities in *monomials*.
(2) One only needs to resolve *minimal* ambiguities. For example, in the $\mathfrak{sl}_2$-case, it is not necessary to resolve the ambiguity $eee \cdot hhh \cdot fff$ - but only to resolve ehf. Let us, in fact, resolve this one now. On the one hand,

$$\begin{aligned} e \cdot hf &\mapsto e(fh - 2f) = ef \cdot h - 2ef \mapsto (fe + h)h - 2(fe + h) \\ &= f \cdot eh + (h^2 - 2fe - 2h) \mapsto f(he - 2e) + h^2 - 2fe - 2h \\ &= fhe - 4fe + h^2 - 2h \end{aligned}$$

and on the other hand,

$$\begin{aligned} eh \cdot f &\mapsto (he - 2e)f = h \cdot ef - 2ef \mapsto h(fe + h) - 2(fe + h) \\ &= hf \cdot e + h^2 - 2fe - 2h \mapsto (fh - 2f)e + h^2 - 2fe - 2h \\ &= fhe - 4fe + h^2 - 2h \end{aligned}$$

Therefore the Diamond Condition holds for $\mathfrak{sl}_2$.

So what are these *minimal ambiguities*? These are only of two types:

(1) *Overlap ambiguities*: These are of the sort $ABC \in \langle X\rangle$, where $W_\sigma = AB$, $W_\tau = BC$ for some $\sigma, \tau \in \mathcal{S}$. The example of ehf above, is an overlap ambiguity.
(2) *Inclusion ambiguities*: These are of the type $W_\sigma = ABC$, $W_\tau = B$ for some $A, B, C \in \langle X\rangle$.

For example, in the case of any Lie algebra (i.e. our second example), there are no inclusion ambiguities, and we shall now verify that the Diamond Condition holds for all overlap ambiguities in general. Modulo the Diamond Lemma, this finishes the proof of the PBW Theorem for Lie algebras.

So let us now assume $i \succ j \succ k \in I$; then we are to resolve the overlap ambiguity $x_i x_j x_k$. The proof is similar to the previous example of $\mathfrak{sl}_2$ above. For simplicity of notation, we rewrite $x_i = x, x_j = y, x_k = z, [x,y] = a_{xy}$ etc. We now compute:

$$\begin{aligned} x \cdot yz &\mapsto x(zy + a_{yz}) = xz \cdot y + xa_{yz} \mapsto (zx + a_{xz})y + xa_{yz} \\ &= z \cdot xy + a_{xz}y + xa_{yz} \mapsto z(yx + a_{xy}) + a_{xz}y + xa_{yz} \\ &= zyx + za_{xy} + a_{xz}y + xa_{yz} \end{aligned}$$

and on the other hand, we have

$$\begin{aligned} xy \cdot z &\mapsto (yx + a_{xy})z = y \cdot xz + a_{xy}z \mapsto y(zx + a_{xz}) + a_{xy}z \\ &= yz \cdot x + ya_{xz} + a_{xy}z \mapsto (zy + a_{yz})x + ya_{xz} + a_{xy}z \\ &= zyx + a_{yz}x + ya_{xz} + a_{xy}z \end{aligned}$$

Therefore both reduction sequences give the same answer if and only if the remaining terms cancel each other, i.e. if and only if we have

$$[z, a_{xy}] + [a_{xz}, y] + [x, a_{yz}] = 0$$

which is precisely the Jacobi identity. Hence we are done, if the Diamond Lemma holds. □

0.7. **Grand Finale - The Diamond Lemma.** So we are now ready to say what the Diamond Lemma says. Simply put: in the above setup ($R = k\langle X\rangle/I$ being the ring in question, and $\mathcal{S}$ the set of reductions), suppose there exists a semigroup partial order compatible with $\mathcal{S}$ - that *also* satisfies the Descending Chain Condition as well as the Diamond Condition. Then the following are equivalent:

(1) All ambiguities are resolvable.
(2) All *minimal* ambiguities are resolvable.
(3) R has a k-basis given by $\langle X\rangle_{\text{irr}}$, the set of irreducible words.

For example, the PBW Theorem for Lie algebras now follows, merely by checking the two types of minimal ambiguities. Scratch that, only one type - the overlap type.

0.8. **Afterthoughts - Another approach.** A second approach to proving the PBW Theorem is as follows: given the reduction system, one notes that $\langle X\rangle_{\text{irr}}$ clearly spans R, and the only thing to prove is linear independence. To show this, define some suitable R-module M, and explicitly write down the action of each element of $\langle X\rangle_{\text{irr}}$ on M. For instance, $M = R$, the regular representation (under left multiplication).

Given such an M, one can show using these explicit formulae, that the elements of $\langle X\rangle_{\text{irr}}$ are indeed linearly independent, when the DCC and Diamond Condition hold. But this is very specific to each of the algebras in question (as is the Diamond Lemma, for that matter). For instance, for $R = \mathfrak{U}(\mathfrak{sl}_2)$, we would define $M = R$, with, for instance, e acting on f via

$T_e(f) := fe + h$ in $\mathfrak{U}(\mathfrak{sl}_2)$, etc. Thus, we see that the two approaches (this one, and the one using the Diamond Lemma) will have essentially the same computations as one another, as far as considering minimal ambiguities goes.

Yet another approach, in the case of *Koszul algebras*, was given by Braverman and Gaitsgory, using Hochschild cohomology, the Bar resolution, and the theory of deformations and obstructions.

References

[Be] G. Bergman, *The diamond lemma for ring theory*, Advances in Mathematics **29** (1978) **2**, 178–218.