1 Intro

Start with an example: consider the following game, a solitaire. Take a finite graph $G$ with $n$ vertices and a function $e : V(G) \rightarrow \mathbb{R}$ which gives values to the vertices of $G$. A move consist of the following steps:

1. Pick any vertex $v^*$ such that $e(v^*) < 0$,
2. Modify $e$ as follows: $\tilde{e}(v^*) = -e(v^*)$ and $\tilde{e}(v) = e(v) - e(v^*)$ for every vertex $v$ adjacent to $v^*$.
3. Repeat with $\tilde{e}$ as the new evaluation function $e$.

The player wins if after a finite number of moves every vertex has non-negative evaluation.

Suppose that I start with some graph $G$ and evaluation function $e$ and I win in $N$ moves. If I give you the same starting position, are you gonna win too even if you don’t know my moves? Can you win in less moves? In more moves?

Claim 1. Suppose the pair $(G, e)$ admits a win in $N$ moves. Then every legal sequence of moves from $(G, e)$ leads to a win in $N$ moves.

Proof. The proof is by induction on $N$, and is obviously true for $N = 0$ (won position) or $N = 1$ (only one vertex has negative evaluation, so there’s only one legal move which is known to be winning).

Suppose then that for every $k < N$, if a pair $(G, \bar{e})$ admits a win in $k$ moves then every legal sequence of moves from $(G, \bar{e})$ leads to a win in $k$ moves. Assume that $(G, e)$ admits a win in $N$ moves starting with the move at the vertex $v_i$, executed by the player $A$. Take any other legal move at $v_j$ executed by the player $D$, we need to show that after this move, $(G, e_j)$ admits a win in $N - 1$ moves.

Note that what we can describe our claim in a very figurative way: let $\Gamma$ be the directed graph whose vertices are evaluations on $G$ (i.e. ordered real $n$-uples $(e_1, \ldots, e_n)$) and where $e$ and $e'$ are linked by an edge going from $e$ to $e'$ if there is a move at some vertex $v_i$ that brings $e$ to $e'$. Then we’re saying that $\Gamma$ is a poset with the 'diamond condition':

```
(1)
```

Kiddie Talk - The Diamond Lemma and its applications

April 22, 2013
Now suppose two players \( B \) and \( C \) join the game. We should distinguish two cases according to the respective position of the vertices \( v_i \) and \( v_j \).

If \( v_i \) and \( v_j \) are not adjacent, then moving first at \( v_i \) and then at \( v_j \) or first at \( v_j \) and then at \( v_i \) leads us in the same configuration \( e_{i,j} \). Suppose that, starting from \( e \), \( C \) moves \( v_i \) then \( v_j \) and \( D \) moves \( v_j \) then \( v_i \), so they both lands in the position \( e_{ij} \).

\( A \) and \( B \) both executed \( v_i \) as the first move, and we know that \( A \) wins in \( N - 1 \) moves after this, so by the induction hypothesis, also \( B \) wins in \( N - 1 \) moves after \( v_i \), and (always by induction hypothesis) \( B \) wins in \( N - 2 \) moves after \( v_j \), where the configuration is \( e_{ij} \). By induction hypothesis between \( B \) and \( C \), \( C \) wins in \( N - 2 \) moves from the position \( e_{ij} \), which means \( C \) wins in \( N - 1 \) moves from the position after the move \( v_j \). But this is the same move \( D \) executed, so by induction hypothesis between \( C \) and \( D \), \( D \) is winning in \( N - 1 \) moves after \( v_j \), i.e. \( D \) wins in \( N \) moves if \( B \) executes the move \( v_j \).

Suppose now \( v_i \) and \( v_j \) are adjacent. Then \( B \) moves at \( v_i \), then \( v_j \) (which is negative at \( e_i \), as its evaluation is \( e(v_j) + e(v_i) \)) then at \( v_i \) again (which is negative at \( e_{i,j} \), as its evaluation is \( e(v_j) \)). Similarly, \( C \) moves at \( v_j \), then \( v_i \), then \( v_j \) again. One verifies that the configurations \( e_{i,j,i} \) and \( e_{j,i,j} \) coincide. Now, after moving at \( v_i \) \( A \) is winning in \( N - 1 \) moves, thus also \( B \) is winning in \( N - 1 \) moves, by induction hypothesis. Therefore at \( e_{i,j,i} \) the player \( B \) is winning in \( N - 3 \) moves, but this is also the configuration obtained by \( C \) at \( e_{j,i,j} \), so after moving at \( v_j \), \( C \) is winning in \( N - 1 \) moves. Then by induction between \( C \) and \( D \), also \( D \) is winning in \( N - 1 \) moves after moving at \( v_j \), so \( D \) wins in \( N \) moves too starting from the initial configuration.

\( \square \)

**Remark.** It is important to have quite clear the proof above, because it gives a very basic sketch of the general proof of the diamond lemma. We are given an initial configuration \( e \) and two different moves which leads to different configuration \( e_1, e_2 \); then we try to construct two sequences of moves starting from \( e_1 \) and \( e_2 \) which eventually join up in some configuration \( f \). Moreover, we have a partial ordering on the configuration and we make sure that \( f < e \), by making sure that every move goes from a 'larger' configuration to a 'smaller' configuration.

## 2 The Diamond Lemma

The Diamond Lemma is, roughly said, a theoretical tool which answers the following questions: given an associative algebra \( A \) defined in terms of generators and relations, is it possible to define a notion of "canonical form" for elements of \( A \)? How is this defined? Which properties must it satisfy?

Let \( X \) be any set, denote with \( \langle X \rangle \) the free monoid on \( X \) with the operation of juxtaposition and with \( k\langle X \rangle \) the free associative \( k \)-algebra on \( X \). We will call *monomials* the elements of \( \langle X \rangle \) and *polynomials* the elements of \( k\langle X \rangle \).

**Definition 1** (Reduction system). A set \( S \) of pairs of the form

\[
\sigma = (W_\sigma, f_\sigma) \text{ for some } W_\sigma \in \langle X \rangle, f_\sigma \in k\langle X \rangle
\]
is said to be a reduction system. For our purposes, we will assume \( W_\sigma \neq f_\sigma \).

**Definition 2** (Reductions). Fix \( A, B \in \langle X \rangle \) monomials and a pair \( \sigma \in S \). We call a reduction the \( k \)-endomorphism of \( k\langle X \rangle \) denoted by \( r_{A\sigma B} \) which sends the monomial \( AW_\sigma B \) to the polynomial \( Af_\sigma B \) and fixes every other monomial.

A reduction \( r_{A\sigma B} \) acts trivially on \( a \in k\langle X \rangle \) if it fixes \( a \), i.e. the monomial \( AW_\sigma B \) does not appear in \( a \). A polynomial is then said to be irreducible if every reduction \( r_{A\sigma B} \) acts trivially on it, for every choices of \( A, B \in \langle X \rangle \) and \( \sigma \in S \).

**Fact 2.** The set \( k\langle X \rangle_{irr} \) of irreducible polynomials is in fact a \( k \)-submodule of \( k\langle X \rangle \).

The questions which now arise spontaneous are: what happens when we start to apply reductions to a polynomial? Does it matter which reduction we apply first? Are we ever going to end, or can we keep applying reductions infinitely many times and never get an irreducible element?

**Definition 3.** Let \( a \in k\langle X \rangle \). A finite sequence of reductions \( r_1, \ldots, r_n \) is said to be final on \( a \) if \( r_n \circ \ldots \circ r_1(a) \) is irreducible.

The polynomial \( a \) is said to be reduction-finite if every infinite sequence of reductions \( (r_n) \) ultimately stabilizes (at an irreducible element).

**Fact 3.** Let \( a \) be reduction-finite. Then every sequence of reductions \( (r_n) \) such that \( r_n \) acts non-trivially on \( r_{n-1} \ldots r_1(a) \) is finite, and is thus a final sequence on \( a \). Furthermore, the reduction-finite elements form a \( k \)-submodule of \( k\langle X \rangle \).

**Definition 4.** A reduction-finite polynomial \( a \) is said to be reduction-unique if every final sequence on \( a \) gives the same irreducible elements, which we denote by \( r_s(a) \in k\langle X \rangle_{irr} \).

**Remark.** We will see that in fact \( r_s \) is a morphism into the irreducible elements, once we restrict the domain to reduction-unique elements.

Let’s now make formal the notion of ambiguity, that is, what to do when we have two (or more) different reductions we could apply on a monomial.

**Definition 5.** A 5-tuple \( (\sigma, \tau, A, B, C) \) with \( \sigma, \tau \in S \) and \( A, B, C \) non-trivial monomials is said to be a overlap ambiguity if \( W_\sigma = AB \) and \( W_\tau = BC \), while it is a inclusion ambiguity if \( W_\sigma = B \), \( W_\tau = ABC \).

In both cases, we have two possible reductions we could start with, and once we applied one we lose the possibility to apply the other. Thus it is not clear if and why applying one or the other should lead to the same irreducible element.

**Definition 6.** An overlap ambiguity is said to be resolvable if there are composition of reductions \( r \) and \( r' \) such that

\[
r(f_\sigma C) = r'(Af_\tau),
\]

that is, the diamond condition is satisfied. Similarly, an inclusion ambiguity is resolvable if there are compositions of reductions \( r \) and \( r' \) such that

\[
r(Af_\sigma C) = r'(f_\tau).
\]
Now that we have defined the reductions and the possible ambiguities we may encounter during the reduction process, we should define an order on $kX$, because only by making it a poset we can somehow formalize the idea of minimal elements (which will correspond to our canonical representatives for a polynomial).

**Definition 7.** A semigroup partial ordering on $X$ is a partial order on $X$ such that

$$B \leq B' \Rightarrow ABC \leq AB'C \quad \forall A, C \in \langle X \rangle.$$  

This is said to be compatible with the reduction system $S$ if for every $\sigma \in S$, $f_\sigma$ is a linear combination of monomials each $< W_\sigma$.

**Remark.** If the semigroup partial ordering $\leq$ is compatible with $S$, after we apply a reduction $r$ to a polynomial $a$, there is no monomial of $r(a)$ strictly larger than any monomial of $a$. This is a fundamental observation!

We also can link the partial ordering with the ambiguities. Let $I$ be the two-sided ideal of $k\langle X \rangle$ generated by

$$I = (W_\sigma - f_\sigma) \text{ as } \sigma \text{ varies in } S.$$  

Given a partial ordering $\leq$ compatible with $S$, for every $A \in \langle X \rangle$ we define the $k$-submodule

$$I_A = \langle B(W_\sigma - f_\sigma)C \rangle \text{ for every } B, C \in \langle X \rangle, \sigma \in S \text{ such that } BW_\sigma C < A.$$  

**Definition 8.** An overlap ambiguity is said to be resolvable relative to $\leq$ if $f_\sigma C - Af_\tau \in I_{ABC}$, that is $f_\sigma C - Af_\tau$ is a linear combination of terms like $D(W_\nu - f_\nu)E$ where $DW_\nu E < ABC$. Similarly, an inclusive ambiguity is resolvable relative to $\leq$ if $Af_\sigma C - f_\tau \in I_{ABC}$.

**Proposition 4.** Let $\leq$ be a partial ordering compatible with $S$. Then every resolvable ambiguity is resolvable relative to $\leq$.

**Proof.** Let $f$ be any polynomial and $r_{A\sigma C}$ a non-trivial reduction on it. A very useful fact is the following: if $g = r_{A\sigma C}(f)$, then

$$f - g = cA(W_\sigma - f_\sigma)C.$$  

Let now $(\sigma, \tau, A, B, C)$ be a resolvable overlap ambiguity, thus we know there are reductions $r = r_1 \ldots r_n$ and $r' = r'_1 \ldots r'_m$ such that $r(f_\sigma C) = r'(Af_\tau)$, where $r_i = r_{A_i \sigma_i C_i}$, $r'_j = R_{A'_j \sigma'_j C'_j}$. Hence we have

$$f_\sigma C = f_n \overset{r_1}{\rightarrow} f_{n-1} \ldots \overset{r'_1}{\rightarrow} f_0,$$

$$Af_\tau = g_m \overset{r'_m}{\rightarrow} g_{m-1} \ldots \overset{r'_1}{\rightarrow} g_0$$

so that by the fact

$$\sum_{h} f_h - f_{h-1} = \sum_{h} f_h - r(f_h) = \sum_{h} c_h A_h (W_\sigma_h - f_\sigma_h) C_h,$$

$$\sum_{k} g_k - g_{k-1} = \sum_{k} f_k - r(f_k) = \sum_{k} c_k A'_k (W_\sigma'_k - f_\sigma'_k) C'_k.$$  

As $\leq$ is compatible with $S$, $f_\sigma C$ and $Af_\tau$ are linear combinations of monomials each $< W_\sigma C = ABC = AW_\tau$, and so are $f_h$ and $g_k$. Among such monomials there are $A_h W_\sigma_h C_h$ and $A'_k W_\sigma'_k C'_k$. Thus

$$f_\sigma C - Af_\tau = \sum_{h} c_h A_h (W_\sigma_h - f_\sigma_h) C_h - \sum_{k} c_k A'_k (W_\sigma'_k - f_\sigma'_k) C'_k \in I_{ABC}$$

so the ambiguity is resolvable relative to $\leq$.

The proof for the inclusive ambiguity works in a completely analogue way. □
We’re now almost ready to introduce Bergman’s original formulation of the Diamond Lemma.

**Lemma 5.** Let $S$ be a reduction system on the free associative $k$-algebra $k\langle X \rangle$ and $\preceq$ a partial order on $\langle X \rangle$ compatible with $S$. Suppose moreover that $\preceq$ satisfies the descending chain condition. Then every element of $k\langle X \rangle$ is reduction-finite.

**Proof.** Let $N$ be the set of elements non reduction-finite. Arguing by contradiction, suppose it is empty, then by the DCC it has a minimal element $M_0$. $M_0$ is not reduction-finite, thus some non-trivial reduction $r_{A\sigma B}$ acts on it, and as $\preceq$ is compatible with $S$, the monomials of $Af_{\sigma}B$ are $< M_0$ and thus reduction-finite, by minimality of $M_0$. This gives a contradiction, hence $N = \emptyset$ and the claim is proved.

**Theorem 6** (Diamond Lemma). Let $S$ be a reduction system on the free associative $k$-algebra $k\langle X \rangle$. Let $\preceq$ be a partial order on $\langle X \rangle$ compatible with $S$ and satisfying the DCC. Then the following conditions are equivalent:

1. Every ambiguity of $S$ is resolvable.
2. Every ambiguity of $S$ is resolvable relative to $\preceq$.
3. Every element of $k\langle X \rangle$ is reduction-unique.
4. A choice of representatives in $k\langle X \rangle$ for the element of the algebra

$$R = k\langle X \rangle/I$$

where $I = (W_{\sigma} - f_{\sigma})$ is a two-sided ideal, is given by the submodule $k\langle X \rangle_{\text{irr}}$ generated by the irreducible monomials in $\langle X \rangle$.

**Proof.**

3 The Poincare-Birkhoff-Witt theorem

Let $k$ be a commutative ring with 2, 3 invertible elements. We consider a Lie algebra $\mathfrak{g}$ which, as a $k$-module, is free over a set $X$.

Consider its tensor algebra

$$T(\mathfrak{g}) = \bigoplus_{n \geq 0} \mathfrak{g}^\otimes n.$$ 

and its ideal

$$I = \langle x \otimes y - y \otimes x - [x, y] \rangle$$

as $x, y$ vary in $\mathfrak{g}$.

Note that $I$ is not a homogeneous ideal.

**Definition 9.** The **universal enveloping algebra** of $\mathfrak{g}$ is defined as

$$U(\mathfrak{g}) = T(\mathfrak{g})/I$$

and it respects the following universal property:

Given the canonical embedding $i : \mathfrak{g} \hookrightarrow T(\mathfrak{g})$ and the projection on the quotient $\pi : T(\mathfrak{g}) \twoheadrightarrow U(\mathfrak{g})$, define the map

$$\tau = \pi \circ i : \mathfrak{g} \rightarrow U(\mathfrak{g})$$

Then for every associative Lie algebra $A$, and map of Lie algebras $f : \mathfrak{g} \rightarrow B$, there exists a unique map of associative Lie algebras $\tilde{f} : U(\mathfrak{g}) \rightarrow B$ such that $f = \tilde{f} \circ \tau$. 

5
Remark. The associative free \( k \)-algebras \( k \langle X \rangle \) and \( T(\mathfrak{g}) \) are isomorphic, thus we can identify

\[
k \langle X \rangle \ni x_1 \ldots x_n \cong x_1 \otimes \ldots \otimes x_n \in T(\mathfrak{g})
\]

for any \( x_1, \ldots, x_n \in X \).

**Proposition 7.** Let \( \leq \) be a total order on \( X \). Then we have

\[
I = \langle xy - yx - [x, y] \rangle \text{ such that } x, y \in X, x < y.
\]

**Proof.** Let \( I' = \langle xy - yx - [x, y] \rangle \) such that \( x, y \in X, x < y \). Clearly \( I' \subseteq I \). Suppose now \( y < x \) are elements of \( X \), then

\[
yx - xy - [y, x] = xy - yx + [x, y] = -(xy - yx - [x, y]) \in I',
\]

hence the elements \( ab - ba - [a, b] \in I' \) for every \( a, b \in X \). But this extends linearly to \( k \langle X \rangle \), thus \( I' \supseteq I \) and the claim is proved. \( \square \)

We are finally ready to prove the Poincare’-Birkhoff-Witt theorem for Lie algebras defined over a ring.

**Theorem 8.** Let \( \mathfrak{g} \) be a Lie algebra, which is a free \( k \)-module with basis \( X = \{x_i\} \). Let \( \leq \) be a total order on \( X \). Then the universal enveloping algebra \( U(\mathfrak{g}) \) is a free module with basis \( 1 \cup \{x_1 \ldots x_n\} \) such that \( x_1 \leq \ldots \leq x_n \).

**Proof.** Define a reduction system \( S \) on \( k \langle X \rangle \) as

\[
\sigma_{xy} = (yx, xy - [x, y]) \text{ if } x < y, x, y \in X.
\]

The ideal generated by the relations is then

\[
\langle W_\sigma - f_\sigma \rangle = \langle yx - xy + [x, y] \rangle = \langle (xy - yx - [x, y]) \rangle = I.
\]

Moreover it is clear that in the canonical map \( \pi : k \langle X \rangle \cong T(\mathfrak{g}) \longrightarrow U(\mathfrak{g}) \), the image of submodule of irreducible elements \( k \langle X \rangle_{irr} \) has as generators exactly the basis given by the claim.

We want to show that every element of \( k \langle X \rangle \) is reduction-finite, and we will use the lemma previous to the Diamond Lemma: so we have to show that there exists a partial order on \( k \langle X \rangle \) compatible with \( S \) which respects the DCC.

Define the **misordering index** of a monomial \( A = x_1 \ldots x_n \in \langle X \rangle \) as the number \( i_A \) of pairs \((i, j)\) such that \( i < j \) but \( x_i > x_j \). Roughly speaking, this tells us how many reductions we will have to apply. Define moreover \( l_A \) as the length of \( A \).

Consider now the following relation on \( \langle X \rangle \): \( A < B \) if \( l_A < l_B \) or \( l_A = l_B \) and \( A, B \) are given by the same elements of the base \( X \), but \( i_A < i_B \). This is clearly a partial order, and it satisfies the antisymmetric property as well as the transitive one.

It is clear that this partial order is compatible with \( S \), because for every \( x < y \) we have \( i_{yx} = 1 > 0 = i(xy) \) and \( l_{xy} = 2 \), while \( [x, y] \in \mathfrak{g} \) and thus is given by a finite linear combinations of elements of \( X \), each of which is a monomial of length 1. To prove that \( < \) is a partial order on \( X \), we have to check that

\[
B < B' \Rightarrow ABC < AB'C \quad \forall A, C \in \langle X \rangle.
\]

If \( l_B < l_{B'} \), this is obvious, otherwise \( l_B = L_{B'} \), \( B' \) is given by a permutation of the same elements of \( X \) which forms \( B \) and \( i_B < i_{B'} \). In that case, one checks all the possibilities for unordered pairs
(i, j) in both ABC and AB′C and finds out that \( i_{ABC} < i_{AB′C} \).

Now we want to show that \( < \) satisfies the DCC on \( \langle X \rangle \). Let \( A_1 > ... > A_n > ... \)
a descending chain. By the way the order is defined, we either have \( l_{A_k} > l_{A_{k+1}} \) or \( l_{A_k} = l_{A_{k+1}} \) but \( i_{A_k} > i_{A_{k+1}} \). We prove DCC by induction on \( n = l_{A_1} \): if \( n = 1 \) it is clear, as the length cannot decrease anymore, so the each step the index \( i_{A_k} \) is decreasing, and in at most \( i_{A_1} \) step the chain stabilizes. Suppose we proved DCC for any chain with an element having length at most \( n - 1 \) and take \( l_{A_1} = n \). Either \( l_{A_2} < n \) and we are done, or \( i_{A_2} < i_{A_1} \). Thus in at most \( i_{A_1} \) steps, we must have an element of length strictly smaller than \( n = l_{A_1} \) and by induction we are done.

In particular, we can apply a lemma above and conclude that every element of \( k\langle X \rangle \) is reduction-finite.

Let now \( a = \sum_i a_i x_i, b = \sum_j b_j y_j \in g \), we have

\[
ab - ba - [a, b] = \sum_{i,j} a_i b_j (x_i y_j - y_j x_i - [x_i, y_j])
\]

and we have

\[
(x_i y_j - y_j x_i - [x_i, y_j]) = (W_{\sigma_{y_j} x_i} - f_{\sigma_{y_j} x_i}) \text{ if } x_i > y_j,
\]

\[
(x_i y_j - y_j x_i - [x_i, y_j]) = -(W_{\sigma_{x_i} y_j} - f_{\sigma_{x_i} y_j}) \text{ if } x_i < y_j.
\]

In particular, for every monomial \( C \) with \( l_C > 2 \), we have

\[
ab - ba - [a, b] \in I_C
\]

as this is sum of terms of the type \( W_{\sigma} - f_{\sigma} \) and \( l_{W_{\sigma}} = 2 < l_C \).

Now we want to show that every ambiguity is resolvable relative to \( < \), then it will be enough to apply the Diamond Lemma \((2) \Rightarrow (4)\) to get the claim.

We have an ambiguity when in a monomial we have terms like \( ... zyx ... \) with \( z > y > x \), or in our formal notation the ambiguities are all of the form \((\sigma_{zy}, \sigma_{yz}, z, y, x)\) for \( z > y > x \). We want to show that

\[
r_{1\sigma_{zy}}(zyx) - r_{2\sigma_{yz}}(zyx) \in I_{zyx}.
\]

Consider first

\[
r_{1\sigma_{zy}}(zyx) = yzx - [y, z]x,
\]

applying subsequent reductions to make it an irreducible element is equivalent to summing elements of \( I_{zyx} \), because \( zyx \) is the maximal monomial of length 3 on \( z, y, x \) without repetitions. Similarly for the second term, so one gets

\[
r_{1\sigma_{zy}}(zyx) - r_{2\sigma_{yz}}(zyx) =
\]

\[
= (x[y, z] - [y, z]x) + ([x, z]y - y[x, z]) + (z[x, y] - [x, y]z) + i
\]

for some \( i \in I_{zyx} \). Up to summing other elements of \( I_{zyx} \), this reduces to the Jacobi identity, which sums to zero, hence \( r_{1\sigma_{zy}}(zyx) - r_{2\sigma_{yz}}(zyx) \in I_{zyx} \).

\( \square \)