**The Reciprocal of 1+ab+aabb+aaabbb+... for NON-COMMUTING a and b, Catalan numbers and non commutative quadratic equations**

**Section 1**

Our goal is to find an inverse of the series $\sum_{n\geq 0} a^n b^n$ where $a$ and $b$ are non-commuting variables. The answer to this question is given by the following theorem.

Let $a, b, x$ be (completely!) **non-commuting** variables ("indeterminates"). Define a sequence of polynomials $d_n(a,b,x)$ ( $n \geq 1$) recursively as follows:

$$d_1(a,b,x) = 1 \quad , \quad (1a)$$

$$d_n(a,b,x) = d_{n-1}(a,b,x)x + \sum_{k=2}^{n-1} d_{n-k}(a,b,x)\, a\, d_k(a,b,x)\, b \quad (n \geq 2) \quad .(1b)$$

Also define the sequence of polynomials $c_n(a,b,x)$ as follows:

$$c_n(a,b,x) = a\, d_n(a,b,x)\, b \quad (n \geq 1) \quad .$$

**Theorem 1**:

$$1 - \sum_{n=1}^{\infty} c_n(a,b,ab-ba) \;=\; \left(\sum_{n\geq 0} a^n b^n\right)^{-1} \quad .$$

It follows immediately that the number of monomials in $a, b$ and $x$ in the polynomial $d_n(a,b,x)$ is the $(n-1)$-th Catalan number. We will give an algebraic and a combinatorial proof of the theorem.

The simplest algebraic proof was given by C. Reutenauer. It is based on two lemmas.

**Lemma 2**: Let $S$ be a formal series in $a$ and $b$ such that $S = 1 + aSb$. The inverse of $S$ is given by series $1 - aDb$ where $D$ satisfies the equation

$$D = 1 + D(x - ab) + DaDb \qquad (2)$$

and $x = ab - ba$.

**Proof**: We are looking for the inverse of $S$ in the form $1 - C$ where $C = aDb$.

We have

$$CS = (1 - S^{-1})S = S - 1 = aSb.$$

Hence

$$C(1 + aSb) = aSb,$$

$$C + CaSb = aSb,$$

$$aDb + aDbaSb = aSb.$$

So,

$$D + DbaS = S$$

and

$$D(1 + baS) = S$$

or

$$D(S^{-1} + ba) = 1.$$

It implies that

$$D(1 - C + ba) = 1$$

and

$$D = 1 + DaDb - Dba$$

which immediately implies equation (2).

**Lemma 3**: Let the degree of indeterminants $a$ and $b$ in equation (2) equals one and the degree of $x$ equals two. Then the solution of equation (2) is given by formula

$$D = \sum_{n \geq 1} d_n(a, b, x)$$

where polynomials $d_n(a, b, x)$ satisfy equations (1).

**Proof**: Note that $D = \sum_{n=1}^{\infty} d_n$ where $d_n = d_n(a, b, x)$ are homogeneous polynomials in $a$ and $b$ of degree $2n - 2$, $n = 1, 2, \ldots$.

The terms of degree 0 and 2 are: $d_1 = 1$ and $d_2 = x$.

Take the term of degree $2n - 2$, $n \geq 3$:

$$d_n = d_{n-1}(x - ab) + \sum_{k=1}^{n-1} d_{n-k} a d_k b = d_{n-1}(x - ab) + d_{n-1}ab + d_1 a d_{n-1} b + \sum_{k=2}^{n-2} d_{n-k} c_k =$$

$$= d_{n-1}x + a d_{n-1} b + \sum_{k=2}^{n-2} d_{n-k} c_k.$$

QED

Let $S = \sum_{n \geq 0} a^n b^n$. Then $S$ satisfies equation $S = 1 + aSb$ and Theorem 1 follows from Lemmas 2 and 3.

**Combinatorial Proof**: Consider the set of *lattice walks* in the 2D rectangular lattice, starting at the origin, $(0,0)$ and ending at $(n-1,n-1)$, where one can either make a *horizontal* step $(i,j) \rightarrow (i+1,j)$, (weight $a$), a *vertical* step $(i,j) \rightarrow (i,j+1)$, (weight $b$) or a diagonal step $(i,j) \rightarrow (i+1,j+1)$, (weight $x$), always staying in the region $i \geq j$, and where you can neither have a horizontal step followed immediately by a vertical step, nor a vertical step followed immediately by a horizontal step. In other words, you may never venture to the region $i < j$, and you can have neither the Hebrew letter Nun (alias the mirror-image of the Latin letter $L$) nor the Greek letter $\Gamma$ when you draw the path on the plane. The weight of a path is the product (in order!) of the weights of the individual steps.

For example, when $n = 2$ the only possible path is $(0,0) \rightarrow (1,1)$, whose weight is $x$.

When $n = 3$ we have two paths. The path $(0,0) \rightarrow (1,1) \rightarrow (2,2)$ whose weight is $x^2$ and the path $(0,0) \rightarrow (1,0) \rightarrow (2,1) \rightarrow (2,2)$ whose weight is $axb$.

When $n = 4$ we have five paths:

The path $(0,0) \rightarrow (1,1) \rightarrow (2,2) \rightarrow (3,3)$ whose weight is $x^3$,

the path $(0,0) \rightarrow (1,0) \rightarrow (2,1) \rightarrow (3,2) \rightarrow (3,3)$ whose weight is $ax^2b$,

the path $(0,0) \rightarrow (1,0) \rightarrow (2,1) \rightarrow (2,2) \rightarrow (3,3)$ whose weight is $axbx$,

the path $(0,0) \rightarrow (1,1) \rightarrow (2,1) \rightarrow (3,2) \rightarrow (3,3)$ whose weight is $xaxb$, and

the path $(0,0) \rightarrow (1,0) \rightarrow (2,0) \rightarrow (3,1) \rightarrow (3,2) \rightarrow (3,3)$ whose weight is $a^2xb^2$.

It is very well-known, and rather easy to see, that the number of such paths are given by the Catalan numbers $C(n-1)$, `http://oeis.org/A000108` .

We claim that the *weight-enumerator* of the set of such walks equals $d_n(a,b,x)$. Indeed, since the walk ends on the diagonal, at the point $(n-1,n-1)$, the last step must be either a diagonal step

$$(n-2,n-2) \rightarrow (n-1,n-1) \quad ,$$

whose weight-enumerator, by the inductive hypothesis is $d_{n-1}(a,b,x)x$, or else let $k$ be the smallest integer such that the walk passed through $(n-k-1,n-k-1)$ (i.e. the penultimate encounter with the diagonal). Note that $k$ can be anything between 2 and $n-1$. The weight-enumerator of the set of paths from $(0,0)$ to $(n-k-1,n-k-1)$ is $d_{n-k}(a,b,x)$, and the weight-enumerator of the set of paths from $(n-k-1,n-k-1)$ to $(n-1,n-1)$ that never touch the diagonal, is $ad_k(a,b,x)b$. So the weight-enumerator is $d_{n-k}(a,b,x)\, a\, d_k(a,b,x)b$ giving the above recurrence for $d_n(a,b,x)$.

It follows that $c_n(a,b,x) = ad_n(a,b,x)b$ is the weight-enumerator of all paths from $(0,0)$ to $(n,n)$ as above with the additional property that except at the beginning $((0,0))$ and the end $((n,n))$

they always stay **strictly** below the diagonal.

Now what does $c_n(a, b, ab - ba)$ weight-enumerate? Now there is a new rule in Manhattan, "no shortcuts", one may not walk diagonally. So every diagonal step $(i, j) \to (i + 1, j + 1)$ must decide whether

to go first horizontally, and then vertically $(i, j) \to (i + 1, j) \to (i + 1, j + 1)$, replacing $x$ by $ab$, or

to go first vertically, and then horizontally $(i, j) \to (i, j + 1) \to (i + 1, j + 1)$, replacing $x$ by $-ba$.

This has to be decided, independently for each of the diagonal steps that formerly had weight $x$. So a path with $r$ diagonal steps gives rise to $2^r$ new paths with sign $(-1)^s$ where $s$ is the number of places where it was decided to go through the second option.

So $c_n(a, b, ab - ba)$ is the weight-enumerators of pairs of paths $[P, K]$ where $P$ is the original path featuring a certain (possibly zero) number of diagonal steps $r$, and $K$ is one of its $2^r$ "children", paths with only horizontal and vertical steps, and weight $\pm weight(C)$, where we have a plus-sign if an even number of the $r$ diagonal steps became *vertical-then-horizontal* (i.e. $ba$) and a minus-sign otherwise.

As we look at the weights of the children $K$ sometimes we have the same path coming from different parents. Let's call a pair $[P, K]$ a *bad* if the path $C$ has a "$ba$" *strictly-under* the diagonal, i.e. a "vertical step followed by a horizontal step" that does not touch the diagonal. Write $K$ as $K = w_1(ba)^s w_2$ where $w_1$ does not have any sub-diagonal $ba$'s and $s$ is as large as possible. Then the parent must be either of the form $P = W_1 x^s W_2$ where the $x^s$ corresponds to the $(ba)^s$, or of the form $P' = W_1 b x^{s-1} a W_2$. In the former case attach $[W_1 x^s W_2, K]$ to $[W_1 b x^{s-1} a W_2, K]$ and in the latter case vice-versa. This is a weight-preserving and **sign-reversing** involution among the bad pairs, so they all kill each other.

It remains to weight-enumerate the *good pairs*. It is easy to see that the good pairs are pairs $[P, K]$ where $K$ has the form $K = a^{i_1} b^{i_1} a^{i_2} b^{i_2} \ldots a^{i_s} b^{i_s}$ for some $s \geq 1$ and integers $i_1, \ldots, i_s \geq 1$ summing up to $n$ (this is called a *composition* of $n$). It is easy to see that for each such $K$, (coming from a good pair $[P, K]$)there can only be **one** possible *parent* $P$. The sign of a good pair

$$[P, a^{i_1} b^{i_1} a^{i_2} b^{i_2} \ldots a^{i_s} b^{i_s}] \quad ,$$

is $(-1)^{s-1}$, since it touches the diagonal $s - 1$ times, and each of these touching points came from an $x$ that was turned into $-ba$.

So $1 - \sum_{n=1}^{\infty} c_n(a, b, ab - ba)$ turned out to be the sum of all the weights of compositions (vectors of positive integers) $(i_1, \ldots, i_s)$ with the weight $(-1)^s a^{i_1} b^{i_1} \cdots a^{i_s} b^{i_s}$ over *all* compositions, but the same is true of

$$\left( \sum_{n \geq 0} a^n b^n \right)^{-1} = \left( 1 + \sum_{n \geq 1} a^n b^n \right)^{-1} = 1 + \sum_{s=1}^{\infty} (-1)^s \left( \sum_{n \geq 1} a^n b^n \right)^s \quad .$$

QED!

**Section 2**

In this section we discuss solutions of noncommutative quadratic equation (2) using quasideterminants. Let $A = (a_{ij})$, $i, j \geq 1$ be a Jacobi matrix, i.e. $a_{ij} = 0$ if $|i - j| > 1$. Set $T = I - A$, where $I$ is the infinite identity matrix. Recall that

$$|T|_{11}^{-1} = 1 + \sum a_{1j_1} a_{j_1 j_2} a_{j_2 j_3} \dots a_{j_k 1}$$

where the sum is taken over all tuples $(j_1, j_2, \dots, j_k)$, $j_1, j_2, \dots, j_k \geq 1$, $k \geq 1$.

Also,

$$|T|_{11} = 1 - a_{11} - \sum a_{1j_1} a_{j_1 j_2} a_{j_2 j_3} \dots a_{j_k 1}$$

where the sum is taken over all tuples $(j_1, j_2, \dots, j_k)$, $j_1, j_2, \dots, j_k > 1$, $k \geq 1$.

Assume that the degree of all diagonal elements $a_{ii}$is two and the degree of all elements $a_{ij}$ such that $i \neq j$ is one. Then

$$|T|_{11}^{-1} = 1 + \sum_{n \geq 1} t_n \qquad (3)$$

where $t_n$ is homogeneous polynomial of degree $2n$ in variables $a_{ij}$.

In particular,

$$t_1 = a_{11} + a_{12} a_{21},$$

$$t_2 = a_{11}^2 + a_{11} a_{12} a_{21} + a_{12} a_{21} a_{11} + a_{12} a_{22} a_{21} + (a_{12} a_{21})^2 + a_{12} a_{23} a_{32} a_{21}.$$

**Problem**: Find the number of monomial terms in $t_n$.

**Proposition 4**: Set $a_{11} = 0$. Then the number of monomials in each each $t_n$ is the $n$-th Catalan number.

**Proof**: ??

Let now $a, x, b$ be formal variables, the degree of $a$ and $b$ is one and the degree of $x$ is two. Set $a_{ii} = x - ab$, $a_{i,i+1} = a, a_{i+1,i} = b$ for all $i$. By the definition of quasideterminants, we have

$$|T|_{11} = 1 - x + ab - a|T|_{11}^{-1} b.$$

Denote $|T|_{11}^{-1}$ by $D$. Then last equation can be written as

$$D^{-1} = 1 - x + ab - aDb$$

or

$$D = D(x - ab) - DaDb$$

which is exactly our equation (2).

**Section 3. Inversion of $1 - aDb$ in general case**.

Let $D = 1 + d_1(a, b, x) + d_2(a, b, x) + \dots$ and polynomials $d_n(a, b, x)$ satisfy equations (1) without any assumptions on $x$. We are looking for the inversion of the series $1 - aDb$ in the form

$$1 + au_1b + au_2b + ...$$

where the degree of $u_n$ is $2n - 2$, $n \geq 1$. Then

$$u_1 = 1,$$

$$u_2 = ba + x,$$

$$u_3 = (ba)^2 + xba + bax + axb + x^2,$$

$$u_4 = (ba)^3 + x(ba)^2 + baxba + (ba)^2x + a^2xb^2 + axb^2a + ba^2xb$$
$$+x^2ba + xbax + bax^2 + ax^2b^2 + axbx + xaxb + x^3,$$

and so on.

**Problem**: How to write a recurrence relations on $u_n$ similar to relations (1). It must imply that the number of terms for $u_n$ is the $n$-th Catalan number. It also must show that if $x = ab - ba$ then $u_n = a^{n-1}b^{n-1}$.

We may set $x = 1$ and get

$$u_1 = 1, \quad u_2 = ba + 1, \quad u_3 = (ba)^2 + 2ba + ab + 1,$$

$$u_4 = (ba)^3 + 3(ba)^2 + ab^2 + ba^2b + a^2b^2 + 3ba + 3ab + 1.$$

**Problem**: How to describe polynoials $u_n$ for this and other specializations? Any relations with known polynomials?