The Reciprocal of 1+ab+aabb+aaabbb+... for NON-COMMUTING a and b

Vladimir RETAKH[1] *and Doron ZEILBERGER*[1]

Proposition: 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:

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

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

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

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

Then

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

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 $(i, j) \to (i+1, j)$, (weight a), a *vertical* step $(i, j) \to (i, j+1)$, (weight b) or a diagonal step $(i, j) \to (i+1, j+1)$, (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 Γ 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 $(0, 0) \to (1, 1)$, whose weight is x.

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

When $n = 4$ we have five paths:

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

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

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

[1] Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA. `[vretakh,zeilberg] at math dot rutgers dot edu` , `http://www.math.rutgers.edu/~vretakh/` , `http://www.math.rutgers.edu/~zeilberg/` . Feb. 22, 2012.

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

the path $(0,0) \to (1,0) \to (2,0) \to (3,1) \to (3,2) \to (3,3)$ 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

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

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

$[P,C]$

where P is the original path featuring a certain (possibly zero) number of diagonal steps r, and C 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 C sometimes we have the same path coming from different parents. Let's call a pair $[P, C]$ a *bad guy* 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 C as $C = w_1(ba)^s w_2$ 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 $[W_1 x^s W_2, C]$ to $[W_1 b x^{s-1} a W_2, C]$ and in the latter case vice-versa. This is a weight-preserving and **sign-reversing** involution among the bad guys, so they all kill each other.

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

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

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) $(i_1, \dots, i_s)$ 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!