Underdiagonal lattice paths with unrestricted steps

Donatella Merlini, D. G. Rogers,* Renzo Sprugnoli, M. Cecilia Verri
Dipartimento di Sistemi e Informatica,
via Lombroso 6/17, Firenze, Italy

Abstract

We use some combinatorial methods to study underdiagonal paths (on the $\mathbf{Z}^2$ lattice) made up of unrestricted steps, i.e., ordered pairs of non-negative integers. We introduce an algorithm which automatically produces some counting generating functions for a large class of these paths. We also give an example of how we use these functions to obtain some specific information on the number $d_{n,k}$ of paths from the origin to a generic point $(n, n-k)$.

Keywords: underdiagonal lattice paths, paths with unrestricted steps, context-free grammars, generating functions, Schützenberger methodology.

1 Introduction

In his interesting paper [3], Gessel gives an algebraic method he calls "factorization of formal Laurent series" to find the generating functions for underdiagonal lattice paths with unrestricted steps (functions f_0 and f_- in his notation) by means of the bivariate generating function of all lattice $\mathbf{Z}^2$'s paths. By *unrestricted steps* we mean ordered pairs (δ, δ') of non-negative integers, and a *path* is a finite sequence of steps starting at the origin; an *underdiagonal path* only contains points (x, y) such that $x \geq y$. Even though literature on lattice paths is extensive, most of it only deals with the steps belonging to some restricted classes. For example, many studies have been made on Dyck paths, but they are only made up of two steps $(0, 1)$ and $(1, 0)$. More in general, researchers seem to prefer treating problems related to "steep steps", i.e., steps (δ, δ') for which $\delta - \delta' \leq 1$, rather than those related to "shallow steps," i.e. steps (δ, δ') for which $\delta - \delta' > 1$. In addition to Gessel's lattice path method, Goldman [4] and Goldman and Sundquist's [5] propose one using a more combinatorial approach, while Labelle's method [6] regards problems involving some kind of restricted steps. The latter's approach consists in starting out with an unambiguous definition of lattice paths by means of a context-free grammar and in then applying Schützenberger's methodology [10] to derive the recurrence relations and generating functions desired.

In the present paper, we apply this method to both underdiagonal paths with unrestricted steps (see Gessel [3]) and underdiagonal paths having privileged access to the main diagonal

*Halewood Cottage, The Green, Croxley Green, United Kingdom, WD3 3HT

i.e., having special steps ending on the main diagonal: lattice paths have often be used as a model for describing the behaviour of a particle walk; in that case, privileged access to the main diagonal corresponds to the fact that the diagonal attracts or rejects the particle. Some meaningful examples of this kind of paths are given in Section 2. We informally describe our main results (see Theorems 2.3 and 2.8) in the following way: a pair (R_A, R_Δ) of finite sets of steps describes a path problem, i.e., the class of underdiagonal lattice paths composed of steps in $R_A \cup R_\Delta$, and precisely: i) R_A is used for steps not ending on the main diagonal, and ii) R_Δ is only used for steps which do end on the main diagonal. Let $d_{n,k}$ be the number of paths ending at the point $(n, n-k)$ (i.e., ending on the diagonal $x - y = k$) and let $D_k(t)$ be the corresponding generating function $\sum_{n=0}^{\infty} d_{n,k}t^n$; in particular, let $D_0(t)$ be the generating function for the paths ending on the main diagonal. It follows that, $D_k(t)$ satisfies a linear recurrence relation:

$$D_k(t) = \Phi(D_{k-1}(t), D_{k-2}(t), \cdots, D_{k-s}(t))$$

whose initial conditions are $D_0(t), D_1(t), \cdots, D_{s-1}(t)$, where s only depends on the maximal difference $|\delta - \delta'|$ for steps in $R_A \cup R_\Delta$.

As far as paths without privileged access to the main diagonal are concerned, we find some of Gessel's results again here and extend them to a more general case. For example, from the context-free formulation of path problems, we can immediately infer that the $D_k(t)$'s (in particular, $D_0(t)$), are algebraic functions. In addition, our approach can be used for deriving some information about the $d_{n,k}$'s ($k > 0$), which are the number of paths not returning to the main diagonal. This is a rather complex problem and is not often treated, especially when shallow steps are involved (see Section 4 for an example of it).

Our paper is organized in the following way: Section 2 contains the definitions, methodology and proofs of our main results. In Section 3, we use the theorems proved in the previous section to introduce an algorithm which starts with the definition of a given lattice path problem and automatically generates the recurrences for the generating functions that solve the problem. Finally, in Section 4, we develop an example to show the difficulties involved in dealing with shallow steps.

2 Definitions and main results

A *step template* is a triple $(\delta, \delta', \kappa)$ where $\delta, \delta' \in \mathbf{N}$ and κ belongs to a (maybe infinite) set of *colours*. A *coloured* $(\delta, \delta', \kappa)$*-step* is a triple $((x, y), (x + \delta, y + \delta'), \kappa)$, where (x, y) and $(x + \delta, y + \delta')$ are two points in $\mathbf{Z}^2$. A *path scheme* R is a finite set of step templates and an *R-path* is a finite sequence $(s_1, s_2, \ldots, s_n)$ of coloured steps such that:

a) $s_1 = ((0, 0), (\delta, \delta'), \kappa)$ with $(\delta, \delta', \kappa) \in R$;

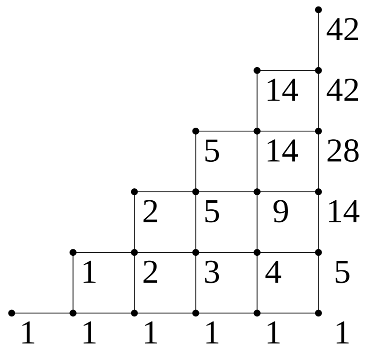
b) $\forall i = 2, 3, \ldots, n$, if $s_i = ((x_i, y_i), (x_i + \delta, y_i + \delta'), \kappa)$ then $x_i \geq y_i$, $x_i + \delta \geq y_i + \delta'$ and $(\delta, \delta', \kappa) \in R$;

c) $\forall i = 2, 3, \ldots, n-1$, if $s_i = ((x_i, y_i), (x_i + \delta, y_i + \delta'), \kappa)$ and $s_{i+1} = ((x_{i+1}, y_{i+1}), (x_{i+1} + \overline{\delta}, y_{i+1} + \overline{\delta}'), \overline{\kappa})$ then $x_i + \delta = x_{i+1}$ and $y_i + \delta' = y_{i+1}$.

The number n is the R-path's *length*. Colours are very useful for representing paths and j equal steps with different colours are algebraically equivalent to a step having *weight* j.

When the set of colours consists in only one element, a path can be described more simply by giving the ordered $(n+1)tuple$ of points $(O, P_1, P_2, \ldots, P_n)$ the path goes through.

Dyck paths constitute a very simple example: their scheme is $R_C = \{(0, 1, black), (1, 0, black)\}$. In Figure 1(a), we illustrate some of these paths. The numbers are usually arranged



(a)

	0	1	2	3	4
0	1				
1	1	1			
2	2	2	1		
3	5	5	3	1	
4	14	14	9	4	1

(b)

Figure 1: Representations for Dyck paths.

in a lower triangular array, as shown in Figure 1(b). The number of paths arriving at the point $(n, n-k)$ is denoted by $d_{n,k}$, which therefore also represents the number contained in the array at row n and column k. The array shown in Figure 1(b) is called the *Catalan triangle*. The Pascal triangle corresponds to the scheme $R_P = \{(1, 0, black), (1, 1, black)\}$ and the Motzkin triangle to $R_M = \{(1, 0, black), (1, 1, black), (1, 2, black)\}$.

We wish to point out that if $((x, y), (x+\delta, y+\delta'), \kappa)$ is a step ending on the main diagonal, i.e., $x + \delta = y + \delta'$, then the step template $(\delta, \delta', \kappa)$ should have $\delta \leq \delta'$. Therefore, if $\Delta(R)$ is R's set of step templates ending on the main diagonal, $\Delta(R)$ is usually different from R. We can generalize our model by defining two sets of step templates: R_A is used for steps *not* ending on the main diagonal, while R_Δ is used for steps which end on the main diagonal. If $\Delta(R_A) = R_\Delta$, we have our original model; if $\Delta(R_A) \neq R_\Delta$, we have some new schemes: a scheme $R = (R_A, R_\Delta)$, for which $\Delta(R_A) \neq R_\Delta$, is said to have *privileged access to the main diagonal.*

Some example are in order at this point. Let us take the artificial scheme $R_A = \{(0, 1, black), (1, 0, black)\}$ and $R_\Delta = \{(0, 1, black), (1, 2, black)\}$. This is a modified Dyck scheme in which the main diagonal attracts particles on the diagonal $x - y = 1$ (see Figure 2(a)). A more interesting example is $R_A = \{(1, 0, black), (1, 1, black), (1, 2, black)\}$ and $R_\Delta = \{(1, 1, black), (1, 2, black), (1, 2, red)\}$. This corresponds to the triangle of trinomial coefficients (see Figure 2(b)). Lastly, if $R = (R_A, R_\Delta)$ is any scheme in the generalized model, its related scheme $R^H = (R_A, \emptyset)$ having privileged access to the main diagonal (because it rejects all the paths except the empty one) is of theoretical interest (see Section 3); the R^H-paths are the R-paths *avoiding the main diagonal.*

We are now going to use the "first and last passage decompositions"method (see Feller [2], p.89) in order to translate a scheme into an unambiguous context-free grammar in which step templates become terminal symbols. The non-terminal symbols are indicated by the names of some path sets, (defined further on). Schützenberger's methodology is then used

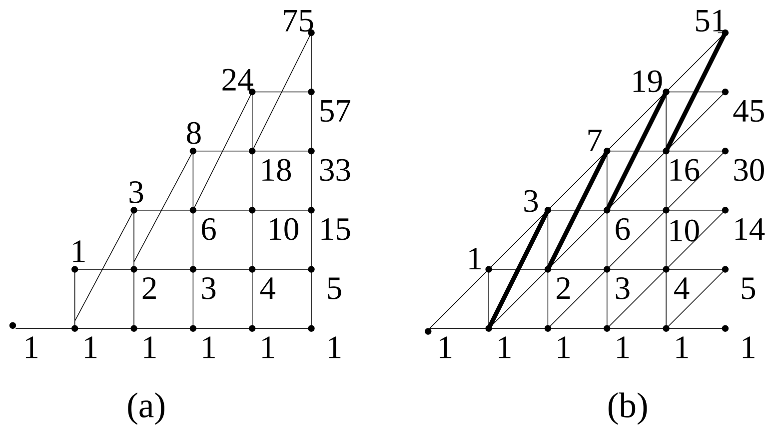


Figure 2: Two schemes illustrating privileged access to the main diagonal.

to obtain some recurrence relations for generating functions that count the number of paths and therefore solve our lattice path problem.

Formally, if $R = (R_A, R_\Delta)$ is a lattice path scheme, we consider the following sets of step templates:

$$S_j = \{(\delta, \delta', \kappa) \in R_A \mid \delta - \delta' = j\}.$$

For simplicity's sake, we start out with the schemes not having privileged access to the main diagonal and prove our results for them. We then state analogous results for our more general model. Let us examine the following sets of R-paths:

- D_k is the set of R-paths that start from the origin $(x_0, y_0) = (0,0)$ and arrive at a point (x_F, y_F) having $x_F - y_F = k$ (i.e., (x_F, y_F) belongs to the diagonal $x - y = k$). In particular, $D_0 = D$ denotes the set of R-paths that reach the main diagonal. More in general, by translation, D_k also denotes the set of R-paths that start from (x_0, y_0), that does not necessarily belong to the main diagonal, arrive at a point (x_F, y_F) having $x_F - y_F \geq x_0 - y_0$ and never go above the diagonal $x - y = x_0 - y_0$;
- $G_s (s \geq 0)$ is the set of R-paths that start from a point (x_0, y_0) having $x_0 - y_0 = k \geq 0$, arrive at the point (x_F, y_F) having $x_F - y_F = k - s \geq 0$ and never go above the diagonal $x - y = x_F - y_F$. In other words, these R-paths climb s units towards the main diagonal without ever going "too high".

It is worth noting that the index k in D_k is only used to relate D_k recursively to some other sets $D_{k'}$ with $k' \leq k$. The index s in G_s takes on the values $0, 1, \ldots, \hat{s}$, where $\hat{s}$ is the maximum value of $\delta' - \delta$ for the steps in R. We now want to find an unambiguous context-free grammar defining D_k and the G_s's. We also need the initial values $D_0, D_1, \cdots, D_r$: D_0 has a specific definition, whilst the other sets are obtained by specializing D_k and their number r depends on the order of the recurrence defining D_k; actually r is the maximum value of $\delta - \delta'$ for the steps in R_Δ.

Lemma 2.1 *Given the scheme R, the set G_s of the R-paths that start from a point (x_0, y_0), arrive at the point (x_F, y_F) having $x_F - y_F = x_0 - y_0 - s$ and never go above the diagonal $x - y = x_F - y_F$, is: $G_0 = D_0$ and for $s > 0$:*

$$G_s ::= D_0 S_{-1} G_{s-1} \left| \begin{array}{c} D_1 S_{-2} G_{s-1} \\ D_0 S_{-2} G_{s-2} \end{array} \right| \begin{array}{c} D_2 S_{-3} G_{s-1} \\ D_1 S_{-3} G_{s-2} \\ D_0 S_{-3} G_{s-3} \end{array} \Bigg| \cdots$$

(This notation only groups some definitions vertically to emphasize their similarity and it is equivalent to the left-to-right BNF notation).

Proof: We immediately deduce $G_0 = D_0$ from the definitions. For $s > 0$, we use the so-called "first passage decomposition": given an R-path in G_s, let us take the first point (x_P, y_P) such that $x_P - y_P < x_0 - y_0$, i.e., the first point at which the path goes above the starting point diagonal $x - y = x_0 - y_0 = k$. The step arriving at (x_P, y_P) should therefore belong to some $S_{-j} (j > 0)$. Let its template be $(\delta, \delta', \kappa)$; the whole path can be divided into three parts, in one and only one way:

i) the path from (x_0, y_0) to $(x_P - \delta, y_P - \delta')$; because, by definition, (x_P, y_P) is the first point above the diagonal $x - y = x_0 - y_0 = k$. This path therefore belongs to D_i, where:

$$i = x_P - \delta - (y_P - \delta') - (x_0 - y_0) = (x_P - y_P) - (x_0 - y_0) - j;$$

ii) the step having template $(\delta, \delta', \kappa) \in S_{-j}$;

iii) the path from (x_P, y_P) to (x_F, y_F); by definition, this path climbs $(x_P - y_P) - (x_F - y_F)$ units towards the main diagonal without ever going above the diagonal $x - y = x_F - y_F$. Therefore, this is a path in $G_{s'}$, with $s' < s$.

Since all the possible values for j are given by the templates in R_A, this relation uniquely determines the corresponding values for k and s'. ■

We are now able to give a grammar for D_0, and this is our first important result:

Theorem 2.2 *Given the scheme R, the set $D_0 = D$ of the R-paths that start at the origin and end on the main diagonal, without ever going above this diagonal, is:*

$$D_0 ::= \epsilon \mid S_0 D_0 \mid S_1 G_1 \mid S_2 G_2 \mid \cdots$$

Proof: The empty path has $(x_F, y_F) = (x_0, y_0) = (0, 0)$ and obviously satisfies the theorem's condition. If the path is not empty, then it should begin with some step having template $(\delta, \delta', \kappa) \in S_j$ $(j \geq 0)$. This step goes j units away from the diagonal's starting point, and, therefore, it should be followed by a path which recuperates these j units without ever going above the diagonal's starting point; this path belongs to G_j by definition. ■

After determining the starting relation for D_0, we go on to find an expression for D_k $(k > 0)$:

Theorem 2.3 *Given a scheme R, the set D_k of the R-paths that start from the origin $(x_0, y_0) = (0, 0)$, arrive at the point (x_F, y_F), with $x_F > y_F$ and always remain on, or below, the main diagonal, is:*

$$D_k ::= D_{k-1}S_1D_0 \left| \begin{array}{c} D_{k-2}S_2D_0 \\ D_{k-1}S_2G_1 \end{array} \right| \begin{array}{c} D_{k-3}S_3D_0 \\ D_{k-2}S_3G_1 \\ D_{k-1}S_3G_2 \end{array} \Bigg| \cdots$$

Proof: We now use the "last passage decomposition" method. Let (x_P, y_P) be the last point at which a path in D_k goes on or below the diagonal $x - y = x_F - y_F$ that starts above it. If the step arriving at (x_P, y_P) has template $(\delta, \delta', \kappa)$ we have $\delta - \delta' = j > 0$ and the whole path can be divided into the following three parts:

i) the path from (x_0, y_0) to $(x_P - \delta, y_P - \delta')$; since the latter point is above the diagonal $x - y = x_F - y_F$, the path is in D_{k-i}, where:

$$k - i = x_P - \delta - (y_P - \delta') - (x_0 - y_0) = (x_P - y_P) - (x_0 - y_0) - j;$$

ii) the step with template $(\delta, \delta', \kappa) \in S_j$;

iii) the path from (x_P, y_P) to (x_F, y_F); since (x_P, y_P) is on, or below, the diagonal $(x - y) = (x_F - y_F)$ and the path never goes above this diagonal, it belongs to G_s, where $s = (x_P - y_P) - (x_F - y_F)$. The decomposition $D_{k-i}S_jG_s$ is obviously unique and we should obtain:

$$k = k - i + j - s \qquad \text{or} \qquad j = i + s.$$

Since the values of j are given by the templates in R, this relation determines the possible values of i and s.

■

Theorems 2.2 and 2.3 define D_k and D_0 in terms of the G_s's. The latter can be eliminated by Lemma 2.1 and we therefore obtain a "recurrence relation" for D_k which depends on some "initial conditions" $D_0, D_1, \cdots, D_r$. The expressions for $D_1, D_2, \cdots, D_r$ are obtained from D_k by setting $D_p = \emptyset$ for $p < 0$. In the next section, we show how these relations can be translated into actual recurrence relations for generating functions.

When we examine a scheme R having privileged access to the main diagonal, we have to consider how a path behaves when it touches the diagonal, because it is not like the other points in $\mathbf{Z}^2$. The sets D_k and G_s become more specialized and should be supported by some auxiliary sets, which we call E_k and G_s^Δ. These sets coincide with the previous one when R does not have privileged access to the main diagonal:

- D_k is the set of R-paths that start from the origin $(x_0, y_0) = (0, 0)$ and arrive at a point (x_F, y_F) having $x_F - y_F = k$. More in general, by translation, D_k also denotes the set of R-paths starting from (x_0, y_0) on the main diagonal and having the same characteristics as the previous ones;

- E_k is the set of R-paths that start from a point (x_0, y_0) not belonging to the main diagonal, arrive at a point (x_F, y_F) having $(x_F - y_F) - (x_0 - y_0) = k$, and never go above the diagonal $x - y = x_0 - y_0$.

- $G_s(s \geq 0)$ is the set of R-paths that start from a point (x_0, y_0) having $x_0 - y_0 = k > 0$ (i.e., a point that does not belong to the main diagonal), arrive at the point (x_F, y_F) having $x_F - y_F = k - s > 0$, and never go above the diagonal $x - y = x_F - y_F$. These paths climb s units towards the main diagonal without ever going "too high";

- $G_s^\Delta(s \geq 0)$ is the set of R-paths that start from a point (x_0, y_0) having $x_0 - y_0 = s$ and arrive at a point (x_F, y_F) belonging to the main diagonal, without ever going above that diagonal.

The following results are stated without proof but the reasoning based on the first and last passage decompositions is analogous to the one previously discussed.

Lemma 2.4 *Given the scheme $R = (R_A, R_\Delta)$, the set G_s of the R-paths that start from a point (x_0, y_0) not belonging to the main diagonal, arrive at the point (x_F, y_F) having $x_F - y_F = x_0 - y_0 - s \neq 0$ and never go above the diagonal $x - y = x_F - y_F$, is: $G_0 = E_0$, and for $s > 0$:*

$$G_s ::= E_0 S_{-1} G_{s-1} \left| \begin{array}{l} E_1 S_{-2} G_{s-1} \\ E_0 S_{-2} G_{s-2} \end{array} \right| \begin{array}{l} E_2 S_{-3} G_{s-1} \\ E_1 S_{-3} G_{s-2} \\ E_0 S_{-3} G_{s-3} \end{array} \Bigg| \cdots$$

■

For paths arriving on the main diagonal we should also consider the following sets of steps:

$$S_j^\Delta = \{(\delta, \delta', \kappa) \in R_\Delta \mid \delta - \delta' = j\}.$$

For $j > 0$, we should have $S_j^\Delta = \emptyset$ and when the scheme has unprivileged access to the main diagonal, $S_j^\Delta = S_j, \forall j \leq 0$.

Lemma 2.5 *Given the scheme $R = (R_A, R_\Delta)$, the set G_s^Δ of the R-paths that start from a point (x_0, y_0) having $x_0 - y_0 = s$ and arrive at a point (x_F, y_F) on the main diagonal, is: $G_0^\Delta = D_0$ and for $s > 0$:*

$$G_s^\Delta ::= E_0 S_{-1}^\Delta G_{s-1}^\Delta \left| \begin{array}{l} E_1 S_{-2} G_{s-1}^\Delta \\ E_0 S_{-2}^\Delta G_{s-2}^\Delta \end{array} \right| \begin{array}{l} E_2 S_{-3} G_{s-1}^\Delta \\ E_1 S_{-3} G_{s-2}^\Delta \\ E_0 S_{-3}^\Delta G_{s-3}^\Delta \end{array} \Bigg| \cdots$$

■

Lemma 2.6 *Given the scheme $R = (R_A, R_\Delta)$, the set E_0 of the R-paths that start from a point (x_0, y_0) not belonging to the main diagonal, arrive at the point (x_F, y_F) on the diagonal $x_F - y_F = x_0 - y_0$ and never go above this diagonal, is:*

$$E_0 ::= \epsilon \mid S_0 E_0 \mid S_1 G_1 \mid S_2 G_2 \mid \cdots$$

Analogously, the set D_0 of the R-paths that start from the origin and end on the main diagonal (without ever going above it) is:

$$D_0 ::= \epsilon \mid S_0 D_0 \mid S_1 G_1^{\Delta} \mid S_2 G_2^{\Delta} \mid \cdots$$

■

Lemma 2.7 *Given a scheme $R = (R_A, R_\Delta)$, the set E_k of the R-paths that start from a point (x_0, y_0) not belonging to the main diagonal, arrive at the point (x_F, y_F) having $x_F - y_F = x_0 - y_0 + k$ and never go above the diagonal $x - y = x_0 - y_0$, is:*

$$E_k ::= E_{k-1} S_1 E_0 \left| \begin{array}{c} E_{k-2} S_2 E_0 \\ E_{k-1} S_2 G_1 \end{array} \right| \begin{array}{c} E_{k-3} S_3 E_0 \\ E_{k-2} S_3 G_1 \\ E_{k-1} S_3 G_2 \end{array} \Bigg| \cdots$$

■

Theorem 2.8 *Given a scheme $R = (R_A, R_\Delta)$, the set D_k of the R-paths that start from the origin $(x_0, y_0) = (0, 0)$, arrive at the point (x_F, y_F), and always stay on, or below, the main diagonal, is:*

$$D_k ::= D_{k-1} S_1 E_0 \left| \begin{array}{c} D_{k-2} S_2 E_0 \\ D_{k-1} S_2 G_1 \end{array} \right| \begin{array}{c} D_{k-3} S_3 E_0 \\ D_{k-2} S_3 G_1 \\ D_{k-1} S_3 G_2 \end{array} \Bigg| \cdots$$

Although the definition of E_k is similar to D_k's their use is quite different because only E_k is needed to obtain $E_1, E_2, \cdots, E_r$. As in the previous model, the G_s's and G_s^{Δ}'s can be eliminated and what remains is the "recurrence" for D_k, which is defined in terms of $D_{k'}$'s with $k' < k$ and the initial conditions $D_0, D_1, \cdots, D_r, E_0, E_1, \cdots, E_r$. In the next section, we go on to treat generating functions' actual recurrence relations and initial conditions.

3 Generating functions and algorithms

The theorems and lemmas illustrated in the previous section can now be used to generate the grammar which recursively defines the set of paths for a given model $R = (R_A, R_\Delta)$. This scheme includes also paths not having privileged access to the main diagonal and so we only study it from now on. The grammar's terminal symbols are R's steps and the symbols D_k and $S_j, D_s, E_s, G_s, G_s^{\Delta}$ (where $s \leq \max\{|\delta - \delta'| \text{ in } R's \text{ steps}\}$) are its non-terminal symbols. The symbols D_k and D_s, E_s are the grammar's most important terms and their definition allows us to determine the recurrence relations for the paths generated by R. By Schützenberger's methodology, we are able to translate the grammar into a set of recursive functional expressions which we derive generating functions $D_k(t)$, $k \in \mathbf{N}$ from. Since for every k the set of steps involved in D_k's definition is finite, every D_k is actually a context-free language and, according to a well-known result [4, 5, 10], the corresponding generating function is algebraic (see also Gessel [3]).

Let us now illustrate the algorithm with a simple but non-trivial example: $R_A = \{(0,1, black), (1,0, black), (3,1, black)\}$ and $R_\Delta = \{(0,1, black)\}$, i.e., a scheme not having privileged access to the main diagonal. From now on, we ignore the latter property in order to describe the various steps of the generating algorithm more clearly.

1) We begin by determining the sets S_j and S_j^Δ and we have:

$$S_1 = \{(1,0, black)\} = \{a\} \qquad S_{-1} = S_{-1}^\Delta \{(0,1, black)\} = \{b\}$$
$$S_2 = \{(3,1, black)\} = \{c\};$$

we denote the three steps by a, b, c to simplify our notations.

2) We now determine the productions for D_k and D_0 by using Lemma 2.6 and Theorem 2.8. In our example, we have:

$$D_0 ::= \epsilon \mid aG_1^\Delta \mid cG_2^\Delta$$
$$D_k ::= D_{k-1}aE_0 \mid D_{k-2}cE_0 \mid D_{k-1}cG_1.$$

This shows that the recurrence relation has order 2 and we need to determine the expressions for E_0, G_1, G_1^Δ, G_2^Δ (as previously observed, in the present case $G_1 = G_1^\Delta$, $G_2 = G_2^\Delta$ and $E_0 = D_0$ but we ignore these identities for clarity's sake).

3) We determine the productions for E_0 by Lemma 2.6:

$$E_0 = \epsilon \mid aG_1 \mid cG_2;$$

this means that we also have to determine G_2,.

4) We determine the productions for G_s and G_s^Δ (s as required); in our example we have $s = 1, 2$ and we find:

$$\begin{aligned} G_0 &= E_0, & G_0^\Delta &= D_0, \\ G_1 &= E_0 b G_0, & G_1^\Delta &= E_0 b G_0^\Delta, \\ G_2 &= E_0 b G_1, & G_2^\Delta &= E_0 b G_1^\Delta. \end{aligned}$$

5) We go on to determine the productions for E_s, G_s, G_s^Δ recursively according to what is generated in the previous step. Since s is limited by the maximal difference $|\delta - \delta'|$ in R's steps, this process eventually ends. In our example, this step is not required.

6) At this point, the grammar is complete except for the initial conditions regarding D_k's recurrence. Since we need as many initial conditions as the order r of the recurrence, and we already know D_0, we should set k to the values from 1 to $r-1$. However $D_1, \cdots, D_{r-1}$ can be found by specializing D_k, this is done by setting $D_j = \emptyset$, $\forall j < 0$. In our example, $r = 2$ and we immediately find:

$$D_1 ::= D_0 a E_0 \mid D_0 c G_1.$$

Since we use the productions for D_k, no new non-terminal symbol is generated.

Our grammar is now complete and can be simplified by standard methods. The final set of productions is:

$$\begin{array}{rcl} D_k & ::= & D_{k-1}aE_0 \mid D_{k-2}cE_0 \mid D_{k-1}cE_0bE_0 \\ D_1 & ::= & D_0aE_0 \mid D_0cE_0bE_0 \\ D_0 & ::= & \epsilon \mid aE_0bD_0 \mid cE_0bE_0bD_0 \\ E_0 & ::= & \epsilon \mid aE_0bE_0 \mid cE_0bE_0bE_0. \end{array}$$

We apply Schützenberger's method to obtain the recurrence relations from the grammar. This method consists in a homomorphism Φ from the grammar to the formal power series' algebra, defined in the following way: first the symbols $::=$ and $\mid$ are transformed into $=$ and $+$; then

$$\Phi(A) = A(t)$$

for every non-terminal A character;

$$\Phi((\delta, \delta', \kappa)) = t^{\delta}$$

if $(\delta, \delta', \kappa)$ is a terminal character, and $\Phi(\epsilon) = 1$. We obtain the following set of functional relations:

$$\begin{array}{rcl} D_k(t) & = & tE_0(t)D_{k-1}(t) + t^3E_0(t)^2D_{k-1}(t) + t^3E_0(t)D_{k-2}(t) \\ D_1(t) & = & tE_0(t)D_0(t) + t^3E_0(t)^2D_0(t) \\ D_0(t) & = & 1 + tE_0(t)D_0(t) + t^3E_0(t)^2D_0(t) \\ E_0(t) & = & 1 + tE_0(t)^2 + t^3E_0(t)^3. \end{array}$$

Finally, we obtain $E_0 = D_0$, and, consequently, $E_0(t) = D_0(t)$. The final solution is therefore:

$$\begin{array}{rcl} D_k(t) & = & tD_0(t)D_{k-1}(t) + t^3D_0(t)^2D_{k-1}(t) + t^3D_0(t)D_{k-2}(t) \\ D_0(t) & = & 1 + tD_0(t)^2 + t^3D_0(t)^3 \\ D_1(t) & = & tD_0(t)^2 + t^3D_0(t)^3. \end{array}$$

In Section 4, we make a detailed description of this recurrence and report the actual values of the $d_{n,k}$ elements in the resulting lower triangular array $\{d_{n,k}\}_{n,k\in\mathbf{N}}$, by counting the number of lattice paths from the origin to the point $(n, n-k)$.

We wish to point out that the bivariate generating function $D(t,w)$ for the complete set of underdiagonal R-paths can be easily obtained from the previous recurrence relation. The function $D(t,w)$ depends on $D_0(t)$, which we denote by $d(t)$ for the simplicity's sake. It is worth noting that $d(t)$ can be found explicitly by solving the third-degree equation that defines it. The reader can use either Maple or Mathematica to solve the problem. By shifting the recurrence relation, we have:

$$D_{k+2}(t) = td(t)D_{k+1}(t) + t^3d(t)^2D_{k+1}(t) + t^3d(t)D_k(t);$$

since this recurrence holds for every $k \in \mathbf{N}$, we can go on to the generating function $D(t,w)$:

$$\frac{D(t,w) - d(t) - D_1(t)w}{w^2} = (td(t) + t^2d(t)^2)\frac{D(t,w) - d(t)}{w} + t^3d(t)D(t,w)$$

and this equation can be easily solved as follows:

$$D(t,w) = \frac{d(t)(1+tw+twd(t))}{1-twd(t)-t^3wd(t)^2-t^2w^2d(t)}.$$

We conclude this section with some considerations on a particular class of schemes in our model. These schemes have been widely treated in current literature (see Raney [9]). They only contain the steps in S_j having $j \leq 1$, and we assume that always have at least one step in S_1. We call them Raney or Riordan schemes and they are characterized by two generating functions $(d(t), h(t))$, as we are now going to see.

Let us begin with the following definition: given a lattice path scheme $R = (R_A, R_\Delta)$, R's *associated scheme* is $R^H = (R_\Delta, \emptyset)$, which has the same set R_A of steps arriving at a point not belonging to the main diagonal and an empty set of templates for the steps that reach the main diagonal. By convention, the set of paths reaching the diagonal $x - y = k$ for the associated scheme are denoted by H_k, instead of D_k. Since R^H does not contain any template for the steps that touch the main diagonal, no R^H-path touches the main diagonal other than at its starting point $(x_0, y_0) = (0, 0)$. Otherwise, the R^H-paths coincide with the R-paths and, as a result, the R^H-paths are called the *R-paths avoiding the main diagonal.* Set H_0 is only made up of the empty path and so we are interested in the H_k's for $k > 0$. There is an important relationship between H_k and D_k, given by the following:

Theorem 3.1 *Given a lattice path problem $R = (R_A, R_\Delta)$ and its associated scheme $R^H = (R_A, \emptyset)$, then for every $k > 0$ we have:*

$$D_k = D_0 H_k.$$

Proof: Let us consider an R-path ending at a point (x_F, y_F) on the diagonal $x - y = k \neq 0$ and let (x_P, y_P) be the last point which the path touches the main diagonal at. The path is divided into two parts:

i) the path from the origin to (x_P, y_P) : it can be $(x_P, y_P) = (0, 0)$, but, at any rate, it is a path in D_0;

ii) the path from (x_P, y_P) to (x_F, y_F) : it is empty if $(x_P, y_P) = (x_F, y_F)$ or, by definition, it avoids the main diagonal; at any rate, it is a path in H_k.

Since this decomposition is unique, the theorem follows from it. ■

As far as Raney-Riordan schemes are concerned, we prove the following:

Theorem 3.2 *Let $R = (R_A, R_\Delta)$ be a Riordan scheme and let $R^H = (R_A, \emptyset)$ be its associated scheme. If $d(t)$ is the generating function of R's main diagonal and $h(t)$ is the generating function of the diagonal $x - y = 1$ in R^H, then the generating function $d_k(t)$ of the diagonal $x - y = k$ is:*

$$d_k(t) = d(t)(th(t))^k. \tag{3.1}$$

Proof: It is worth noting that, in general, if R and R' are two schemes having $R_A = R'_A$, then $E_0 = E'_0$ because they do not depend on R_Δ and R'_Δ, as the results in the previous section show. This is particularly true for an R scheme and its R^H associated scheme. By Theorem 2.8, we now have $D_k = D_{k-1}S_1E_0$ for a Riordan scheme, and, according to the preceding remarks, $H_k = H_{k-1}S_1E_0$. For $k = 1$, the latter relation reduces to $H_1 = S_1E_0$, and we therefore have $D_k = D_{k-1}H_1$. The generating function is equivalent to $D_k(t) = D_{k-1}(t)tH_1(t)$, which can be iterated to give $D_k(t) = D_0(t)(tH_1(t))^k$. If we set $D_0(t) = d(t)$ and $H_1(t) = h(t)$, we immediately obtain the result desired. ■

This theorem provides us with a simple way for dealing with Riordan schemes: we only need to compute $d(t) = D_0(t)$ with the original scheme $R = (R_A, R_\Delta)$ and then compute $h(t) = H_1(t)$ by the associated scheme $R^H = (R_A, \emptyset)$ in order to describe the whole array $\{d_{n,k}\}_{n,k\in\mathbf{N}}$. In current literature, any array characterized by two formal power series $(d(t), h(t))$ such that the generating function for column k is given by (3.1) is called a Riordan array (see [11]). Many path properties in a Raney-Riordan scheme can be studied by means of Riordan array theory, and the reader is referred to [8] for further information on the subject.

There is another way of computing the function $h(t)$ without having to deal with the associated scheme explicitly:

Lemma 3.3 *Let $R = (R_A, R_\Delta)$ be a Riordan scheme and let $R^H = (R_A, \emptyset)$ be its associated scheme; then we have:*

$$H_1 = S_1E_0.$$

Proof: By applying Theorem 2.8 to the associated scheme R^H, we obtain:

$$H_1 = H_0S_1E_0 \mid H_0S_2G_1 \mid H_0S_3G_2 \mid \cdots .$$

Since $H_0 = \{\epsilon\}$, $S_j = \emptyset \ \forall j > 1$ the lemma follows. ■

When S_1 only contains a finite number of step templates, $S_1(t)$ is a polynomial and we have $h(t) = S_1(t)E_0(t)$. When the scheme has unprivileged access to the main diagonal, then $h(t) = S_1(t)d(t)$ and, in particular, if S_1 only contains the step template $(1, 0, black)$, then $h(t) = d(t)$. A Riordan array having $d(t) = h(t)$ is said to be a *renewal array* and its elements are computed by the following simplified formula:

$$d_{n,k} = [t^n]d(t)(th(t))^k = [t^{n-k}]d(t)^{k+1}.$$

4 The analysis of a simple case

We conclude by giving a closer look at a particular example, that is the scheme $R = \{(1, 0, black), (0, 1, black), (3, 1, black)\}$ having unprivileged access to the main diagonal. This case can also be studied by Gessel's method [3]. The same functional relation for D_0 can be easily found for the paths that go back to the main diagonal. As far as the other kinds of paths are concerned, our approach obtains recurrence relations which may be better suited for asymptotic analysis. On the other hand, some explicit bivariate generating functions can also be derived.

The formulas found in the previous section can be analyzed further to obtain some more precise information on the number of paths generated by this simple scheme. Let us begin by $d_n = [t^n]d(t)$, the number of paths arriving at the point (n,n) on the main diagonal. If we set $y = y(t) = td(t)$, the formula for $d(t)$ can be written as:

$$\frac{y}{t} = 1 + \frac{y^2}{t} + y^3 \qquad \text{or} \qquad y = t\frac{1+y^3}{1-y}.$$

Since $y(0) = 0$, we can apply the Lagrange inversion formula and obtain an explicit expression for $d_n = [t^{n+1}]y(t)$:

$$d_n = \frac{1}{n+1}[y^n]\left(\frac{1+y^3}{1-y}\right)^n = \frac{1}{n+1}\sum_{k=0}^{n/3}\binom{n+1}{k}\binom{2n-3k}{n-3k}.$$

From the former formula, we can derive the asymptotic expression for d_n by using the method of implicit functions (see Markushevitch [7]) as described in Sprugnoli and Verri [12]. However, we obtain the same expression by using the following version of the same method. This, in turn, can be directly applied to some other cases of lattice path enumeration. Let $\mathcal{F}(t,d) = 1 - d + td^2 + t^3d^3$ be the functional equation defining $d = d(t)$; the dominating singularities of $d(t)$ (i.e., the singularities having the smallest modulus) are among the solutions of the following system:

$$\begin{cases} \mathcal{F}(t,d) = 1 - d + td^2 + t^3d^3 = 0 \\ \mathcal{F}'_d(t,d) = -1 + 2td + 3t^3d^2 = 0 \end{cases}.$$

It is quite simple to obtain the solution (r,s) we are interested in:

$$r = \frac{1}{3}\sqrt{2\sqrt{3}-3} \approx .2270833462, \quad s = \left(\frac{5}{2}\sqrt{3}+\frac{9}{2}\right)\sqrt{2\sqrt{3}-3} - \frac{3}{2}(\sqrt{3}+1) \approx 1.917448161$$

Since $\mathcal{F}''_{dd}(r,s) \neq 0$, the function $d = d(t)$ can be developed around its dominating singularity:

$$d(t) = s + a_1\left(1-\frac{t}{r}\right)^{1/2} + a_2\left(1-\frac{t}{r}\right) + a_3\left(1-\frac{t}{r}\right)^{3/2} + \cdots$$

We can now determine the values of $a_1, a_2, a_3, \ldots$ by feeding this expression into the functional equation $\mathcal{F}(t,d) = 0$ and by equating the coefficients to 0. This gives us a series of linear equations which can be solved in the variables $a_1, a_2, a_3, \ldots$. By using Maple, we obtain:

$$a_1 = -s, \qquad a_2 = \left(\frac{9}{2}\sqrt{3}+8\right)\sqrt{2\sqrt{3}-3} - \frac{7}{2} - 3\sqrt{3} \approx 2.063666450,$$

$$a_3 = -\left(\frac{55}{12}\sqrt{3}+\frac{49}{6}\right)\sqrt{2\sqrt{3}-3} + \frac{37}{12}\sqrt{3} + \frac{7}{2} \approx -2.131200536,$$

$$a_4 = \left(\frac{59}{9}\sqrt{3}+\frac{139}{12}\right)\sqrt{2\sqrt{3}-3} - \frac{41}{9}\sqrt{3} - \frac{67}{12} \approx 2.152649536,$$

$$a_5 = -\left(\frac{1913}{288}\sqrt{3}+\frac{1129}{96}\right)\sqrt{2\sqrt{3}-3} + \frac{449}{96}\sqrt{3} + \frac{535}{96} \approx -2.175658428$$

and, therefore, we have the following asymptotic approximation:

$$d_n \sim \left(a_1 \binom{1/2}{n} + a_3 \binom{3/2}{n} + a_5 \binom{5/2}{n} \right) \left(-\frac{1}{r} \right)^n.$$

The main term of this expression can now be found by applying the well-known asymptotic formula for $\binom{1/2}{n}$:

$$d_n \sim \frac{s}{2n\sqrt{\pi n} r^n}.$$

Let us now go on to $d_{n,k} = [t^n]d_k(t)$, where $d_k(t)$ is given by the recurrence in the formulas used in Example 4.2. The initial conditions are $d_0(t) = d(t)$ and $d_1(t) = td_0(t)(d_0(t) + d_0(t)^2t^2) = d(t) - 1$, by the formula defining $d_0(t)$. Therefore:

$$d_2(t) = t^3 d(t)^2 + td_1(t)h(t) = t^3 d(t)^2 + d(t) - 2 + d(t)^{-1}.$$

This expression can be simplified by using the initial relation in the form of $t^3 d(t)^2 = 1 - td(t) - d(t)^{-1}$:

$$d_2(t) = (1-t)d(t) - 1.$$

The quantity $d(t)^{-1}$ has disappeared here, and we can prove the following result:

Lemma 4.1 *For every value $m \in \mathbf{N}$, we have:*

$$d_m(t) = p_m(t)d(t) - q_m(t),$$

where $p_m(t)$ and $q_m(t)$ are polynomials, such that $q_m(t) = p_{m-1}(t), p_0(t) = 1, q_0(t) = 0$.

Proof: The three functions $d_0(t), d_1(t)$ and $d_2(t)$ correspond to the theorem's initial cases and they allow us to proceed by mathematical induction.

$$d_m(t) = d_{m-1}(t)(1 - d(t)^{-1}) + t^3 d_{m-2}(t)d(t) =$$

$$= (p_{m-1}(t)d(t) - q_{m-1}(t))(1 - d(t)^{-1}) + t^3 d(t)(p_{m-2}(t)d(t) - q_{m-2}(t)) =$$

$$= p_{m-1}(t)d(t) - p_{m-1}(t) - q_{m-1}(t) + q_{m-1}(t)d(t)^{-1} + t^3 p_{m-2}(t)d(t)^2 - t^3 d(t)q_{m-2}(t) =$$

$$= p_{m-1}(t)d(t) - p_{m-1}(t) - q_{m-1}(t) + q_{m-1}(t)d(t)^{-1} + p_{m-2}(t) -$$

$$-tp_{m-2}(t)d(t) - p_{m-2}(t)d(t)^{-1} - t^3 q_{m-2}(t)d(t) =$$

$$= (p_{m-1}(t) - tp_{m-2}(t) - t^3 q_{m-2}(t))d(t) - p_{m-1}(t) - q_{m-1}(t) + p_{m-2}(t) + (q_{m-1}(t) - p_{m-2}(t))d(t)^{-1}.$$

By the induction hypothesis, $q_{m-1}(t) = p_{m-2}(t)$ and, therefore, $d(t)^{-1}$ disappears. We also obtain:

$$q_m(t) = p_{m-1}(t) + q_{m-1}(t) - p_{m-2}(t) = p_{m-1}(t)$$

which proves the final relation. ■

It is obvious that the polynomials $q_m(t)$ are simple corrections which reduce the initial part of the generating functions $d_m(t)$ to 0. As a result, the asymptotic value of $d_{n,k} = [t^n]d_k(t)$ only depends on the product $p_k(t)d(t)$ and the polynomial $p_k(t)$ is definitely analytic at the dominating singularity $t = r$ of $d(t)$. By a well-known result (see Bender [1], Theorem 2), it follows that

$$d_{n,k} \sim p_k(r)[t^n]d(t).$$

Since we already know the asymptotic value of $[t^n]d(t)$, we only have to compute $p_k(r)$ for every $k \in \mathbf{N}$. This is not difficult and we can find a closed form for $p_k(r)$:

Lemma 4.2 *The closed form for the value of $p_k(t)$ computed at $t = r$ is:*

$$p_k(r) = A(k+1)s^{-k} + Bs^{-k} + Cs_1^{-k},$$

where s has the same value as before, $s_1 = -(1+2sr^2)/r^2 \approx -23.22720118$ and A, B, C are the three constants

$$A = \frac{1+2sr^2}{1+3sr^2} \approx .9237432928, \quad B = \frac{sr^2(1+2sr^2)}{1+6sr^2+9s^2r^4} \approx .0704416217,$$

$$C = \frac{s^2r^4}{1+6sr^2+9s^2r^4} \approx .0058150853.$$

Proof: From the last step in the derivation of $d_m(t)$ in Lemma 4.1's proof, we have:

$$p_k(t) = p_{k-1}(t) - tp_{k-2}(t) - t^3q_{k-2}(t) = p_{k-1}(t) - tp_{k-2}(t) - t^3p_{k-3}(t);$$

we therefore obtain the recurrence:

$$p_k(r) = p_{k-1}(r) - rp_{k-2}(r) - r^3p_{k-3}(r),$$

whose initial conditions are $p_0(r) = 1, p_1(r) = 1$ and $p_2(r) = 1 - r$. If we write the recurrence for $k+3$ and examine the generating function $\mathcal{G}\{p_k(r)\} = P(t)$, we have:

$$p_{k+3}(r) = p_{k+2}(r) - rp_{k+1}(r) - r^3p_k(r),$$

$$\frac{P(t) - 1 - t - (1-r)t^2}{t^3} = \frac{P(t) - 1 - t}{t^2} - r\frac{P(t) - 1}{t} - r^3P(t),$$

$$P(t) = \frac{1}{1 - t + rt^2 + r^3t^3}.$$

The denominator's roots can be easily found: they are $t_1 = t_2 = s$ and $t_3 = -(1+2sr^2)/r^2 = s_1$. A partial fraction expansion now gives us:

$$P(t) = \frac{A}{(1-t/s)^2} + \frac{B}{1-t/s} + \frac{C}{1-t/s_1},$$

and we can immediately derive the closed form from it. ∎

Thanks to all the previous results, we can conclude with the asymptotic formula for the generic element $d_{n,k}$, i.e., the number of lattice paths arriving at the point $(n, n-k)$:

Theorem 4.3 *The asymptotic value of $d_{n,k} = [t^n]d_k(t)$ is:*

$$d_{n,k} \sim ((A(k+1) + B)s^{-k} + Cs_1^{-k})d_n \sim \frac{Ak}{2n\sqrt{\pi n}}\frac{1}{s^{k-1}r^n}.$$

Proof: The proof is immediate from the two lemmas above and from the previously-mentioned theorem by Bender. It is not difficult (but very laborious) to find out that this formula is a true approximation when $k = o(\sqrt{n})$. ∎

Aknowledgements

We wish to thank the anonymous referee whose comments helped us in improving the contents and readibility of our paper.

References

[1] E. A. Bender. Asymptotic methods in enumeration. *SIAM Review*, 16:485–515, 1974.

[2] W. Feller. *An introduction to probability theory and its applications.* Wiley, 1950.

[3] I. M. Gessel. A factorization for formal Laurent series and lattice path enumeration. *Journal of Combinatorial Theory, Series A*, 28:321–337, 1980.

[4] J. R. Goldman. Formal languages and enumeration. *Journal of Combinatorial Theory, Series A*, 24:318–338, 1978.

[5] J. R. Goldman and T. Sundquist. Lattice path enumeration by formal schema. *Advances in applied mathematics*, 13:216–251, 1992.

[6] J. Labelle. Languages de Dyck généralisés. *Ann. Sci. Math. Québec*, 17(1):53–64, 1993.

[7] A.I. Markushevich. *Theory of functions of a complex variable.* Prentice-Hall, 1965.

[8] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri. On some alternative characterizations of Riordan arrays. *Canadian Journal of Mathematics*, 49(2):301–320, 1997.

[9] G. N. Raney. Functional composition patterns and power series reversion. *Trans. Amer. Math. Soc*, 94:441–451, 1960.

[10] M. P. Schützenberger. Context-free language and pushdown automata. *Information and Control*, 6:246–264, 1963.

[11] R. Sprugnoli. Riordan arrays and combinatorial sums. *Discrete Mathematics*, 132:267–290, 1994.

[12] R. Sprugnoli and M.C. Verri. Asymptotics for Lagrange inversion. *Pure Mathematics and Applications*, 5(1), 1994.