POLLARD'S RHO ALGORITHM

Bruno Salvy
(Version of January 27, 1997)

Pollard's ρ-method is an efficient technique used to find factors of integers. It is both very efficient and very simple. We show in this worksheet how combstruct and gfun can be used to analyze a realistic combinatorial model of the algorithm and thus derive a probabilistic complexity analysis of this algorithm and variants of it.

```
> with(combstruct): with(gfun):
```

Algorithm and combinatorial model

The algorithm

If N is the integer of which a factor is sought, the basic version of the algorithm is as follows:

Pick up an arbitrary integer a mod N, set $\mathrm{f}(x) = (x^2 + a) \bmod N$
Select at random a number x_0 mod N, set $y_0 = x_0$
Iterate:
$i := i + 1;\ x_i := \mathrm{f}(x_{i-1});\ y_i := \mathrm{f}(\mathrm{f}(y_{i-1}))$
until $\gcd(y_i - x_i, N) \neq 1$.

This is directly translated into the following rough Maple procedure which returns a factor and the number of iterations performed to find this factor:

```
> pollard:=proc(N)
  local rnd, a, f, x, y, i, g;
      rnd:=rand(N); a:=rnd(); x:=rnd(); y:=x;
      for i do
          x:=x^2+a mod N; y:=(y^2+a mod N)^2+a mod N;
  g:=igcd(y-x,N);
          if g<>1 then RETURN(g,i) fi
      od
  end:
```

Here are a few examples (Maple's nextprime routine returns the smallest prime number following its argument).

```
> nextprime(900)*nextprime(20000);
```

$$18149977$$

```
> pollard(");
```

$$907, 31$$

```
> pollard("");
```

$$907, 32$$

```
> pollard(""");
```

$$907, 30$$

```
> nextprime(10^5)*nextprime(10^6);
```

$$100003300009$$

```
> pollard(");
```

$$1000003, 117$$

```
> pollard("");
```

100003, 536

```
> pollard(""");
```

100003, 233

```
> nextprime(10^6)*nextprime(10^7);
```

10000049000057

```
> pollard(");
```

1000003, 792

```
> pollard("");
```

1000003, 390

```
> pollard(""");
```

10000019, 1370

The combinatorial model

The algorithm relies on the structure of the functional graph of the function $f(x) = (x^2 + a) \bmod p$, where p is the smallest prime factor of N. In this graph, the vertices are the integers $0 .. p - 1$ mod p and the directed edges link each vertex to its image by f. Since the number of points is finite, it is not difficult to see that the graph has the structure of a union of connected components, each of these components being constituted of a cycle to which converge trees. Since the polynomial $f(x)$ has degree 2 and p is prime, all the vertices (except a) have in-degree 0 or 2, while they have out-degree 1. The prime factor p is assumed to be large, therefore the special case of the vertex a which has in-degree 1 can be discarded as a first approximation. The combinatorial model is thus expressed by the following combstruct grammar:

```
> G:={fungraph=Set(connected_component),
      connected_component=Cycle(Prod(Z,bintree)),
      bintree=Union(Z,Prod(Z,Set(bintree,card=2)))}:
```

The execution of the algorithm is interpreted on this graph as follows: a random point x_0 of the graph is selected. Then two sequences x_i and y_i of iterates with initial value x_0 are computed, one of them proceeding one step at a time, while the other one proceeds by steps of length 2. Starting from a node of the graph, these two sequences eventually reach a cycle, where they are bound to intersect. This is where the algorithm stops.

Under this model, the average number of steps required by the algorithm is therefore related to two parameters: *(i)* the average distance from a point to its cycle; *(ii)* the average length of the cycles.

Path length in planar binary trees

We start with question *(i)* above: the determination of the average distance from a point to its cycle. We first concentrate on a similar but simpler problem, the computation of the average distance from a node to the root in a planar binary tree. This is related to a classical combinatorial parameter: the *internal path length* of the tree, which is the sum of the distances from each of the nodes to the root.

Binary trees are described by the following grammar:

```
> bin:={bintree=Union(Epsilon,Prod(Z,bintree,bintree))}:
```

The counting sequence given by combstruct[count] is constituted by the well-known Catalan numbers:

```
> seq(count([bintree,bin,unlabelled],size=i),i=0..15);
```

1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, 2674440,
9694845

The following Maple procedure takes as input a binary tree as produced by combstruct using the specification above and returns its internal path length:

```
> ipl:=proc(t)
     if type(t,name) then 0 # the tree is of the form Z or
  Epsilon
     else # the tree is of the form Prod(Z,t1,t2)
        ipl(op(2,t))+ipl(op(3,t))+size(t)-1
     fi
  end:
  size:=proc(t) eval(subs([Prod=`+`,Z=1,Epsilon=0],t)) end:
```

The internal path length is computed using a simple bijection: each of the nodes on a branch from the root to a leaf are counted once for each of the nodes below it. Here is an example on a tree of size 5 generated by combstruct[draw] from the specification above:

```
> T:=draw([bintree,bin,unlabelled],size=5); ipl(T);
```

$$T := \mathrm{Prod}(Z, \mathrm{Prod}(Z, \mathrm{E}, \mathrm{Prod}(Z, \mathrm{Prod}(Z, \mathrm{Prod}(Z, \mathrm{E}, \mathrm{E}), \mathrm{E}), \mathrm{E})), \mathrm{E})$$

$$10$$

We shall compute the average internal path length by first computing the total internal path length, i.e. the sum of the internal path lengths of all the binary trees of size n and then dividing this number by the nth Catalan number. We first start by computing experimentally the first values of these numbers, generating all the trees of fixed sizes for small sizes with combstruct[allstructs], and computing the sum of the internal path lengths of all these trees with the procedure ipl above.

```
> for i to 5 do
  sum_ipl[i]:=`+`(op(map(ipl,allstructs([bintree,bin,unlabelled
  ],size=i)))) od;
```

$$sum\_ipl_1 := 0$$

$$sum\_ipl_2 := 2$$

$$sum\_ipl_3 := 14$$

$$sum\_ipl_4 := 74$$

$$sum\_ipl_5 := 352$$

Combinatorial model for path lengths

The numbers computed above can actually be derived more efficiently, together with many other results related to path lengths using combstruct's ability to deal with marks (atoms of size 0). The idea consists in writing a combinatorial specification such that the number of objects of size n is exactly the sum of the internal path lengths of all the binary trees of size n. The combinatorial technique used here extends to other combinatorial structures and leads to a combstruct-based analysis of Pollard's ρ-algorithm.

The grammar for binary trees is modified to take into account "decorated" binary trees. A binary tree is decorated by putting two marks (An for Ancestor and De for Descendant) on two nodes belonging to the same branch. The number of ways of decorating a binary tree is then exactly its internal path length. Couting the number of decorated trees therefore corresponds to summing the internal path lengths of all binary trees.

```
> bin2:={bintree2=Prod(Z,
              Union(leftright2,Prod(An,leftright1))),
          bintree1=Prod(Z,
              Union(leftright1,Prod(De,bintree,bintree))),
          leftright2=Union(Prod(bintree2,bintree),
                           Prod(bintree,bintree2)),
          leftright1=Union(Prod(bintree1,bintree),
```

```
                          Prod(bintree,bintree1)),
        An=Epsilon, De=Epsilon} union bin:
```

The sequence of cumulated internal path lengths is now derived in less than a second:

```
> seq(count([bintree2,bin2,unlabelled],size=i),i=1..10);
```

$$0, 2, 14, 74, 352, 1588, 6946, 29786, 126008, 527900$$

Here are the 14 decorated trees of size 3:

```
> allstructs([bintree2,bin2,unlabelled],size=3);
```

[Prod(Z, Prod(An, Prod(E, Prod(Z, Prod(De, E, %2))))),
Prod(Z, Prod(An, Prod(Prod(Z, Prod(De, E, %2)), E))),
Prod(Z, Prod(E, Prod(Z, Prod(An, Prod(E, %1))))),
Prod(Z, Prod(An, Prod(%2, %1))),
Prod(Z, Prod(An, Prod(E, Prod(Z, Prod(%1, E))))),
Prod(Z, Prod(E, Prod(Z, Prod(An, Prod(%1, E))))),
Prod(Z, Prod(An, Prod(%1, %2))),
Prod(Z, Prod(An, Prod(Prod(Z, Prod(De, %2, E)), E))),
Prod(Z, Prod(An, Prod(E, Prod(Z, Prod(E, %1))))),
Prod(Z, Prod(An, Prod(E, Prod(Z, Prod(De, %2, E))))),
Prod(Z, Prod(Prod(Z, Prod(An, Prod(%1, E))), E)),
Prod(Z, Prod(Prod(Z, Prod(An, Prod(E, %1))), E)),
Prod(Z, Prod(An, Prod(Prod(Z, Prod(%1, E)), E))),
Prod(Z, Prod(An, Prod(Prod(Z, Prod(E, %1)), E)))]
%1 := Prod(Z, Prod(De, E, E))
%2 := Prod(Z, E, E)

Generating functions

From the grammars describing binary trees and decorated binary trees, the average internal path length can be computed via generating functions. Using combstruct[gfsolve], we first derive the generating functions for the Catalan numbers and for the cumulated path lengths:

```
> F:=subs(gfsolve(bin,unlabelled,z),bintree(z));
```

$$F := \frac{1}{2}\,\frac{1-\sqrt{1-4\,z}}{z}$$

```
> S:=series(F,z,30);
```

$$S := 1 + z + 2\,z^2 + 5\,z^3 + 14\,z^4 + 42\,z^5 + 132\,z^6 + 429\,z^7 + 1430\,z^8 + 4862\,z^9 + 16796\,z^{10} + 58786\,z^{11} + 208012\,z^{12} + 742900\,z^{13} + 2674440\,z^{14} + 9694845\,z^{15} + 35357670\,z^{16} + 129644790\,z^{17} + 477638700\,z^{18} + 1767263190\,z^{19} + 6564120420\,z^{20} + 24466267020\,z^{21} + 91482563640\,z^{22} + 343059613650\,z^{23} + 1289904147324\,z^{24} + 4861946401452\,z^{25} + 18367353072152\,z^{26} + 69533550916004\,z^{27} + 263747951750360\,z^{28} + \mathrm{O}(z^{29})$$

```
> Fpl:=subs(gfsolve(bin2,unlabelled,z),bintree2(z));
```

$$Fpl := 2\,\frac{-\frac{1}{2}\,\frac{1-\sqrt{1-4\,z}}{z}+\frac{3}{2}-\frac{1}{2}\sqrt{1-4\,z}}{-1+4\,z}$$

```
> Spl:=series(Fpl,z,30);
```

$Spl := 2\,z^2 + 14\,z^3 + 74\,z^4 + 352\,z^5 + 1588\,z^6 + 6946\,z^7 + 29786\,z^8 + 126008\,z^9 + 527900\,z^{10} + 2195580\,z^{11} + 9080772\,z^{12} + 37392864\,z^{13} + 153434536\,z^{14} + 627778954\,z^{15} + 2562441466\,z^{16} + 10438340104\,z^{17} + 42449348236\,z^{18} + 172376641924\,z^{19} + 699100282156\,z^{20} + 2832205421824\,z^{21} + 11462854280536\,z^{22} + 46354571222164\,z^{23} + 187311973956004\,z^{24} + 756391980332272\,z^{25} + 3052578734670488\,z^{26} + 12312647334369656\,z^{27} + 49639018139147336\,z^{28} + \mathrm{O}(z^{29})$

The average path length is simply the ratio of these coefficients:

```
> seq(coeff(Spl,z,i)/coeff(S,z,i),i=0..28);
```

$0, 0, 1, \frac{14}{5}, \frac{37}{7}, \frac{176}{21}, \frac{397}{33}, \frac{6946}{429}, \frac{14893}{715}, \frac{63004}{2431}, \frac{131975}{4199}, \frac{1097790}{29393}, \frac{2270193}{52003}, \frac{9348216}{185725}, \frac{19179317}{334305}, \frac{627778954}{9694845}, \frac{1281220733}{17678835}, \frac{5219170052}{64822395}, \frac{10612337059}{119409675}, \frac{86188320962}{883631595}, \frac{174775070539}{1641030105}, \frac{708051355456}{6116566755}, \frac{1432856785067}{11435320455}, \frac{23177285611082}{171529806825}, \frac{46827993489001}{322476036831}, \frac{189097995083068}{1215486600363}, \frac{381572341833811}{2295919134019}, \frac{3078161833592414}{17383387729001}, \frac{6204877267393417}{32968493968795}$

```
> evalf(["]);
```

$[0, 0, 1., 2.800000000, 5.285714286, 8.380952381, 12.03030303, 16.19114219, 20.82937063, 25.91690662, 31.43010241, 37.34868846, 43.65503913, 50.33364383, 57.37071536, 64.75389282, 72.47201148, 80.51492161, 88.87334346, 97.53874969, 106.5032689, 115.7596056, 125.3009735, 135.1210384, 145.2138706, 155.5739035, 166.1958978, 177.0749109, 188.2062697]$

Empirically, these numbers grow slightly faster than linearly with the size of the tree. A closed-form for the average path length can be established rigorously using gfun. The generating functions being algebraic, they satisfy linear differential equations. These can be derived by gfun[holexprtodiffeq]. From these differential equations, a linear recurrence satisfied by the Taylor coefficients is obtained by gfun[diffeqtorec]. It turns out that these recurrences fall into a class for which Maple's rsolve can find a solution.

Here are the differential equations:

```
> deqF:=holexprtodiffeq(F,y(z));
```

$$deqF := 1 + (-1 + 2\,z)\,\mathrm{y}(z) + (-z + 4\,z^2)\left(\frac{\partial}{\partial z}\mathrm{y}(z)\right)$$

```
> deqFpl:=holexprtodiffeq(Fpl,y(z));
```

$deqFpl := \{\,(D^{(2)})(y)(0) = 4,\ \mathrm{y}(0) = 0,\ \mathrm{D}(y)(0) = 0,\ 6\,z^2 + (8\,z^3 - 26\,z^2 + 10\,z - 1)\,\mathrm{y}(z) + (16\,z^4 - 24\,z^3 + 9\,z^2 - z)\left(\frac{\partial}{\partial z}\mathrm{y}(z)\right)\}$

From there, the recurrences follow:

```
> recF:=diffeqtorec(deqF,y(z),u(n));
```

$$recF := \{\,\mathrm{u}(0) = 1,\ (2 + 4\,n)\,\mathrm{u}(n) + (-2 - n)\,\mathrm{u}(n+1)\,\}$$

```
> recFpl:=diffeqtorec(deqFpl,y(z),u(n));
```

$recFpl := \{ (8 + 16\,n)\,\mathrm{u}(n) + (-50 - 24\,n)\,\mathrm{u}(n+1) + (28 + 9\,n)\,\mathrm{u}(n+2) + (-4 - n)\,\mathrm{u}(n+3),$

$$u(2)=2,\ u(0)=0,\ u(1)=0\,\}$$

The recurrence satisfied by the Catalan numbers being linear of order 1, it is obvious that Maple will be able to solve it:

```
> Cat:=rsolve(recF,u(n));
```

$$Cat := \frac{4^n\,\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+2)\sqrt{\pi}}$$

It is however more surprising that rsolve can compute the solution of the 3rd order recurrence satisfied by the cumulated path lengths. This is due to the implementation of the recent algorithm by M. Petkovsek for hypergeometric solutions of linear recurrences.

```
> Pl:=rsolve(recFpl,u(n));
```

$$Pl := -3\,\frac{4^n\left(n+\frac{1}{3}\right)\left(\frac{1}{2}\right)^n 2^n\,\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+2)\sqrt{\pi}}+4^n$$

After some cleaning up, we have thus proved the following.

Proposition. *The average internal path length in a random planar unlabelled binary tree of size n under the uniform model is*

$$\frac{4^n\,(n+1)}{\mathrm{binomial}(2\,n,n)}-3\,n-1.$$

```
> seq(4^i*(i+1)/binomial(2*i,i)-3*i-1,i=1..10);
```

$$0,\ 1,\ \frac{14}{5},\ \frac{37}{7},\ \frac{176}{21},\ \frac{397}{33},\ \frac{6946}{429},\ \frac{14893}{715},\ \frac{63004}{2431},\ \frac{131975}{4199}$$

From there, the asymptotic behaviour is well within the reach of Maple's asympt command:

```
> map(combine,asympt(Pl/Cat,n),exp);
```

$$\frac{\sqrt{\pi}}{\left(\frac{1}{n}\right)^{3/2}}-3\,n+\frac{9}{8}\,\frac{\sqrt{\pi}}{\sqrt{\frac{1}{n}}}-1+\frac{17}{128}\sqrt{\pi}\,\sqrt{\frac{1}{n}}+\frac{3}{1024}\sqrt{\pi}\left(\frac{1}{n}\right)^{3/2}+\mathrm{O}\left(\left(\frac{1}{n}\right)^{5/2}\right)$$

Average distance from a point to a cycle

We now come back to the analysis of Pollard's ρ-algorithm. The analysis of the average distance from a node to its cycle proceeds exactly as in the case of binary trees by decorating the corresponding trees. The grammar is derived from the grammar G above:

```
> G;
```

$$\{\mathit{fungraph}=\mathrm{Set}(\mathit{connected\_component}),$$
$$\mathit{connected\_component}=\mathrm{Cycle}(\mathrm{Prod}(Z,\mathit{bintree})),$$
$$\mathit{bintree}=\mathrm{Union}(Z,\mathrm{Prod}(Z,\mathrm{Set}(\mathit{bintree},\mathit{card}=2)))\}$$

The process of decoration consists in isolating the path between ancestor and descendant in the combinatorial structure. For instance, non-planar binary trees are described by

```
> npbin:={bintree=subs(G,bintree)};
```

$$\mathit{npbin} := \{\mathit{bintree}=\mathrm{Union}(Z,\mathrm{Prod}(Z,\mathrm{Set}(\mathit{bintree},\mathit{card}=2)))\}$$

Here, Set indicates that the respective position of the two subtrees does not count. In the labelled case, we can therefore decide that the decorated branch will always be the leftmost one (instead of considering all the paths leftright1 and leftright2 as in the case of planar binary trees). Thus the decorated grammar in this case is

```
> bin3:={bintree2=Prod(Z,Union(Prod(bintree2,bintree),
```

```
                              Prod(An,bintree1,bintree))),
        bintree1=
            Prod(Z,Union(Prod(bintree1,bintree),
                         Prod(De,Set(bintree,card=2)),De)),
        An=Epsilon,De=Epsilon} union npbin:
```

Again, we can check that the first values coincide with the result produced by computing the internal path length on all the non-planar binary trees:

```
> ipl:=proc(t) if type(t,name) then 0 else
  size(t)-1+ipl(op([2,1],t))+ipl(op([2,2],t)) fi end:
  size:=proc(t) local
  i;eval(subs([Prod=`+`,Set=`+`,seq(i=1,i=indets(t,name))],t))
  end:
> for i to 6 do
  i,`+`(op(map(ipl,allstructs([bintree,npbin,labelled],size=i))
  )) od;
```

$$1, 0$$

$$2, 0$$

$$3, 6$$

$$4, 0$$

$$5, 360$$

$$6, 0$$

```
> seq(count([bintree2,bin3,labelled],size=i),i=1..11);
```

$$0, 0, 6, 0, 360, 0, 36540, 0, 5896800, 0, 1402077600$$

The same process readily extends to the functional graphs involved in Pollard's algorithm by a decomposition of sets and cycles which isolates the marked part.

```
> G1:={fungraph=Prod(connected_component1,
                     Set(connected_component)),
      connected_component1=Prod(Z,
          Union(Prod(An,bintree1),bintree2),
                Sequence(Prod(Z,bintree))),
      connected_component=Cycle(Prod(Z,bintree))}
  union bin3:
```

We first write procedures to compute the distance from nodes to their cycles in the non-decorated graphs so that we can check on the first few values that our grammar for decorated graphs is consistent with the non-decorated one.

```
> iplfg:=proc(g) `+`(op(map(iplcc,g))) end:
  iplcc:=proc(cc) `+`(op(map(iplbt,cc))) end:
  iplbt:=proc(bt) if type(bt,name) then 0 else
  size(op(2,bt))+iplt(op(2,bt)) fi end:
  iplt:=proc(t) if type(t,name) then 0 else
  size(t)-1+iplt(op([2,1],t))+iplt(op([2,2],t)) fi end:
  size:=proc(t) local i;
  eval(subs([Set=`+`,Prod=`+`,seq(i=1,i=indets(t,name))],t))
  end:
```

Here are a few tests:

```
> to 4 do t:=draw([fungraph,G,labelled],size=6);
  print(t,iplfg(t)) od:
```

$$\mathrm{Set}(\mathrm{Cycle}(\mathrm{Prod}(Z_1, Z_5)), \mathrm{Cycle}(\mathrm{Prod}(Z_3, \mathrm{Prod}(Z_4, \mathrm{Set}(Z_2, Z_6))))), 6$$

$$\mathrm{Set}(\mathrm{Cycle}(\mathrm{Prod}(Z_4, Z_1), \mathrm{Prod}(Z_6, \mathrm{Prod}(Z_5, \mathrm{Set}(Z_2, Z_3))))), 6$$

$$\mathrm{Set}(\mathrm{Cycle}(\mathrm{Prod}(Z_5, Z_1)), \mathrm{Cycle}(\mathrm{Prod}(Z_4, \mathrm{Prod}(Z_6, \mathrm{Set}(Z_2, Z_3))))), 6$$

$$\mathrm{Set}(\mathrm{Cycle}(\mathrm{Prod}(Z_5, \mathrm{Prod}(Z_4, \mathrm{Set}(Z_1, \mathrm{Prod}(Z_6, \mathrm{Set}(Z_2, Z_3))))))), 11$$

The number of functional graphs of fixed size grows quite fast:

```
> seq(count([fungraph,G,labelled],size=i),i=1..10);
```

$$0, 2, 0, 36, 0, 1800, 0, 176400, 0, 28576800$$

We can therefore only check the values for sizes 2 and 4 in a reasonable amount of time

```
> map(iplfg,allstructs([fungraph,G,labelled],size=2));
```

$$[1, 1]$$

```
> sort(map(iplfg,allstructs([fungraph,G,labelled],size=4)));
```

$$[2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5]$$

```
> convert(",`+`);
```

$$108$$

On the other hand, here is the counting sequence for decorated graphs:

```
> seq(count([fungraph,G1,labelled],size=i),i=1..10);
```

$$0, 2, 0, 108, 0, 10440, 0, 1638000, 0, 382384800$$

It takes 11 minutes to check that 10440 is also the value we get by generating all the binary functional graphs of size 6.

We now compute the generating functions with combstruct[gfsolve]:

```
> F:=subs(gfsolve(G,labelled,z),fungraph(z));
```

$$F := \frac{1}{\sqrt{-2\,z^2+1}}$$

```
> Fpl:=subs(gfsolve(G1,labelled,z),fungraph(z));
```

$$Fpl := \frac{z\left(-\frac{1}{2}\,\frac{2-2\sqrt{-2\,z^2+1}}{z}+2\,z\right)}{(1-4\,z^2+4\,z^4)\sqrt{-2\,z^2+1}}$$

Again, we obtain closed-forms from these generating functions by applying gfun to find the differential equation they satisfy and from there the recurrence satisfied by their coefficients:

```
> deqF:=holexprtodiffeq(F,y(z));
```

$$deqF := \{2\,z\,\mathrm{y}(z)+(-1+2\,z^2)\left(\frac{\partial}{\partial z}\,\mathrm{y}(z)\right), \mathrm{y}(0)=1\}$$

```
> deqFpl:=holexprtodiffeq(Fpl,y(z));
```

$deqFpl := \{$

$$\mathrm{y}(0)=0, \mathrm{D}(y)(0)=0, (24\,z^5-24\,z^3+6\,z)\,\mathrm{y}(z)+(6\,z^2-1+8\,z^6-12\,z^4)\left(\frac{\partial}{\partial z}\,\mathrm{y}(z)\right)+2\,z$$

$\}$

```
> recF:=diffeqtorec(deqF,y(z),u(n));
```

$$recF := \{\mathrm{u}(0)=1, (2\,n+2)\,\mathrm{u}(n)+(-2-n)\,\mathrm{u}(n+2), \mathrm{u}(1)=0\}$$

```
> recFpl:=diffeqtorec(deqFpl,y(z),u(n));
```

$$recFpl := \{\mathrm{u}(0)=0, \mathrm{u}(1)=0, \mathrm{u}(3)=0, \mathrm{u}(5)=0, \mathrm{u}(2)=1, \mathrm{u}(4)=\frac{9}{2},$$

$$(24+8\,n)\,\mathrm{u}(n)+(-48-12\,n)\,\mathrm{u}(n+2)+(6\,n+30)\,\mathrm{u}(n+4)+(-n-6)\,\mathrm{u}(n+6)\}$$

Unfortunately, due to a weakness in Maple's current version of rsolve, the solutions to these recurrences are not found

```
> rsolve(recF,u(n)),rsolve(recFpl,u(n));
```

$$\mathrm{rsolve}(\{\mathrm{u}(0)=1, (2\,n+2)\,\mathrm{u}(n)+(-2-n)\,\mathrm{u}(n+2), \mathrm{u}(1)=0\}, \mathrm{u}(n)), \mathrm{rsolve}\Big(\{$$

$$u(0)=0,\ u(1)=0,\ u(3)=0,\ u(5)=0,\ u(2)=1,\ u(4)=\frac{9}{2},$$

$$(24+8n)\,u(n)+(-48-12n)\,u(n+2)+(6n+30)\,u(n+4)+(-n-6)\,u(n+6)\},$$

$$u(n)\Big)$$

However, Maple can find the solution if we help it by taking into account the parity of the generating functions:

```
> deqF2:=holexprtodiffeq(subs(z=z^(1/2),F),y(z)):
> deqFpl2:=holexprtodiffeq(subs(z=z^(1/2),normal(Fpl)),y(z)):
> recF2:=diffeqtorec(deqF2,y(z),u(n));
```

$$recF2 := \{u(0)=1,\ (1+2n)\,u(n)+(-n-1)\,u(n+1)\}$$

```
> recFpl2:=diffeqtorec(deqFpl2,y(z),u(n));
```

$$recFpl2 := \{u(0)=0,\ (12+8n)\,u(n)+(-12n-24)\,u(n+1)+(15+6n)\,u(n+2)+(-n-3)\,u(n+3),\ u(1)=1,\ u(2)=\frac{9}{2}\}$$

```
> nb_fg:=subs(n=n/2,rsolve(recF2,u(n)));
```

$$nb\_fg := \frac{2^{(1/2\,n)}\,\Gamma\left(\frac{1}{2}+\frac{1}{2}n\right)}{\Gamma\left(1+\frac{1}{2}n\right)\sqrt{\pi}}$$

```
> tot_pl:=map(simplify,subs(n=n/2,rsolve(recFpl2,u(n))));
```

$$tot\_pl := 2^{(1/2\,n)}+2^{(1/2\,n-1)}\,n-\frac{2^{(1+1/2\,n)}\,\Gamma\left(\frac{3}{2}+\frac{1}{2}n\right)}{\Gamma\left(1+\frac{1}{2}n\right)\sqrt{\pi}}$$

This is summarized by:

Proposition: *The average distance from a point to a cycle in a random binary non-planar functional graph of size n is given by* $\dfrac{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}+2\right)}{n\,\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}-1-\dfrac{1}{n}$.

From there a direct asymptotic expansion follows:

```
> map(simplify,asympt(subs(tot_pl/nb_fg/n),n));
```

$$\frac{1}{4}\frac{\sqrt{2}\sqrt{\pi}}{\sqrt{\frac{1}{n}}}-1+\frac{9}{16}\sqrt{2}\sqrt{\pi}\sqrt{\frac{1}{n}}-\frac{1}{n}+\frac{17}{128}\frac{\sqrt{2}\sqrt{\pi}\sqrt{\frac{1}{n}}}{n}+\frac{3}{512}\frac{\sqrt{2}\sqrt{\pi}\sqrt{\frac{1}{n}}}{n^2}-\frac{181}{8192}\frac{\sqrt{2}\sqrt{\pi}\sqrt{\frac{1}{n}}}{n^3}+O\left(\frac{\sqrt{\frac{1}{n}}}{n^4}\right)$$

```
> as_cost_1:=op(1,"):
```

Thus under our probabilistic model, the first stage of Pollard's ρ-algorithm takes asymptotically $C\sqrt{p}$ steps, where p is the smallest prime factor of the number to be factored and $C = \frac{\sqrt{2\pi}}{4}$.

Average length of the cycles

After both sequences (x_i and y_i) have reached the cycle, the number of steps before the end of Pollard's algorithm is bounded by the length of this cycle. Since this length might be correlated to the number of steps before, it is not a priori sufficient to compute the average length of the cycles in a random graph obeying our grammar. Instead, we modify the decorated graphs so that the Ancestor is now an element of the cycle. The number of decorated graphs of size n is thus the sum for all the nodes of the length of their cycle, summed over all graphs of size n. Here is the new grammar:

```
> G2:=remove(type,G1,identical(connected_component1)=anything)
  union {connected_component1=
        Prod(Z,
             Union(bintree1,Prod(De,bintree)),
             Sequence(Prod(Z,bintree)),
             Prod(Z,An,bintree),
             Sequence(Prod(Z,bintree)))}:
```

The corresponding generating function is again algebraic:

```
> Fplcycle:=factor(normal(subs(gfsolve(G2,labelled,z),fungraph(
  z))));
```

$$Fplcycle := -2\,\frac{z^2\left(-1+\sqrt{-2z^2+1}\right)}{\left(-1+2z^2\right)^2}$$

```
> series(Fplcycle,z,11);
```

$$2z^4+9z^6+29z^8+\frac{325}{4}z^{10}+O(z^{12})$$

Again, a closed form follows for the coefficients:

```
> holexprtodiffeq(subs(z=z^(1/2),Fplcycle)/z,y(z));
```

$$\left\{2+(3+12z^2-12z)\,y(z)+(6z-1+8z^3-12z^2)\left(\frac{\partial}{\partial z}y(z)\right), y(0)=0\right\}$$

```
> diffeqtorec(",y(z),u(n));
```

$$\{u(0)=0,\ u(1)=2,\ u(2)=9,$$
$$(12+8n)\,u(n)+(-12n-24)\,u(n+1)+(15+6n)\,u(n+2)+(-n-3)\,u(n+3)\}$$

```
> rsolve(",u(n));
```

$$2\,2^n-4\,\frac{(2^n)^2\left(\frac{1}{2}\right)^n\Gamma\left(\frac{3}{2}+n\right)}{\Gamma(n+1)\sqrt{\pi}}+2\,2^n\,n$$

```
> T1:=map(simplify,subs(n=n/2-1,")+tot_pl);
```

$$T1 := 2\,2^{(1/2\,n)}-\frac{2^{(1+1/2\,n)}\,\Gamma\left(\frac{1}{2}+\frac{1}{2}n\right)}{\Gamma\left(\frac{1}{2}n\right)\sqrt{\pi}}+2^{(1/2\,n-1)}(n-2)+2^{(1/2\,n-1)}\,n$$

$$-\frac{2^{(1+1/2\,n)}\,\Gamma\left(\frac{3}{2}+\frac{1}{2}n\right)}{\Gamma\left(1+\frac{1}{2}n\right)\sqrt{\pi}}$$

We thus get the following conclusion:

Theorem. *The average number of steps of Pollard's* ρ*-algorithm under the probabilistic model is bounded between* $\frac{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}+2\right)}{n\,\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}-1-\frac{1}{n}$ *and* $\frac{2\sqrt{\pi}\,\Gamma\left(\frac{n}{2}+1\right)}{n\,\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}-2-\frac{1}{n}$*, where n is the smallest prime factor of the integer whose factorization is sought.*

Asymptotically, this number of steps therefore behaves like $C\sqrt{n}$*, with C in* $\frac{\sqrt{2\,\pi}}{4}\,..\,\frac{\sqrt{2\,\pi}}{2}$.

Here is the verification of the second part:

```
> map(simplify,asympt(T1/nb_fg/n,n,3));
```

$$\frac{1}{2}\frac{\sqrt{2}\sqrt{\pi}}{\sqrt{\frac{1}{n}}}-2+\frac{5}{8}\sqrt{2}\sqrt{\pi}\sqrt{\frac{1}{n}}-\frac{1}{n}+\mathrm{O}\left(\frac{\sqrt{\frac{1}{n}}}{n}\right)$$

```
> as_cost_2:=op(1,"):
> evalf("");
```

$$\frac{1.253314137}{\sqrt{\frac{1}{n}}}-2.+1.566642671\sqrt{\frac{1}{n}}-\frac{1.}{n}+\mathrm{O}\left(\frac{\sqrt{\frac{1}{n}}}{n}\right)$$

Since the smallest prime factor of a number N is of order $\mathrm{O}(\sqrt{N})$, it follows that the (arithmetic) complexity of Pollard's algorithm under this model grows in $\mathrm{O}\left(N^{\left(\frac{1}{4}\right)}\right)$.

Extension: polynomials of degree *m*

When the factor p to be found is known a priori to satisfy $p = 1 \bmod m$ for some m known in advance, Pollard conjectured that the number of steps in his algorithm could be reduced by a factor $\sqrt{m-1}$ by using the polynomial $\mathrm{f}(x) = (x^m + a) \bmod N$ instead of $\mathrm{f}(x) = (x^2 + a) \bmod N$. Brent and Pollard used this idea to factor the 8th Fermat number, whose factors are known to be $1 \bmod 2^{(8+2)}$. We know consider Pollard's under the probabilistic model for $m = 6$.

Here is the grammar for functional graphs with m as a parameter:

```
> Gm:={fg=Set(cc),cc=Cycle(Prod(Z,Set(tree,card=m-1))),
       tree=Union(Z,Prod(Z,Set(tree,card=m)))}:
```

Here is the corresponding decorated grammar for the distances to the cycles:

```
> Gm2:={fg1=Prod(cc1,Set(cc)),
        cc1=Prod(Z,Union(Prod(An,tree1),tree2),
                 Set(tree,card=m-2),
                 Sequence(Prod(Z,Set(tree,card=m-1)))),
        tree1=Prod(Z,Union(Prod(tree1,Set(tree,card=m-1)),
                           Prod(De,Set(tree,card=m)),De)),
```

```
        tree2=Prod(Z,Union(tree2,Prod(An,tree1)),
                     Set(tree,card=m-1)),
        An=Epsilon,De=Epsilon} union Gm:
```

And the grammar for the distances to the cycles plus the length of the cycle:

```
> Gm3:={fg2=Prod(cc2,Set(cc)),
        cc2=Prod(Z,Union(Prod(tree1,Set(tree,card=m-2)),
                                 Prod(De,Set(tree,card=m-1))),
                     Sequence(Prod(Z,Set(tree,card=m-1))),
                     Prod(Z,An,Set(tree,card=m-1)),
                     Sequence(Prod(Z,Set(tree,card=m-1))))}
  union Gm2:
```

$m = 2$

We first check on the case $m = 2$ that we recover the generating functions we had before.

```
> sol:=gfsolve(subs(m=2,Gm3),labelled,z);
```

$$sol := \left\{ \mathrm{An}(z) = 1, \mathrm{Z}(z) = z, \mathrm{De}(z) = 1, \mathrm{cc}(z) = \ln\left(\frac{1}{\sqrt{-2z^2+1}}\right), \mathrm{tree}(z) = \frac{1}{2}\frac{\%1}{z},\right.$$
$$\mathrm{fg1}(z) = \frac{z\left(-\frac{1}{2}\frac{\%1}{z}+2z\right)}{(1-4z^2+4z^4)\sqrt{-2z^2+1}}, \mathrm{tree2}(z) = 2\,\frac{-\frac{1}{2}\frac{\%1}{z}+z}{-1+2z^2},$$
$$\mathrm{fg2}(z) = \frac{z^2\,\%1}{(\sqrt{-2z^2+1}-2z^2+z^2\,\%1)\sqrt{-2z^2+1}}, \mathrm{tree1}(z) = \frac{1}{2}\frac{\%1}{z\sqrt{-2z^2+1}},$$
$$\left.\mathrm{cc2}(z) = \frac{z^2\,\%1}{\sqrt{-2z^2+1}-2z^2+z^2\,\%1}, \mathrm{cc1}(z) = \frac{z\left(-\frac{1}{2}\frac{\%1}{z}+2z\right)}{1-4z^2+4z^4}, \mathrm{fg}(z) = \frac{1}{\sqrt{-2z^2+1}}\right\}$$

$$\%1 := 2 - 2\sqrt{-2z^2+1}$$

```
> normal(subs(sol,fg(z))-F);
```

$$0$$

```
> normal(subs(sol,fg1(z))-Fpl);
```

$$0$$

```
> normal(subs(sol,fg2(z))-Fplcycle);
```

$$0$$

$m = 6$

The grammar becomes

```
> G6:=subs(m=6,Gm3);
```

$G6 := \{fg = \mathrm{Set}(cc), fg1 = \mathrm{Prod}(cc1, \mathrm{Set}(cc)), An = \mathrm{E}, De = \mathrm{E},$

$fg2 = \mathrm{Prod}(cc2, \mathrm{Set}(cc)), tree = \mathrm{Union}(Z, \mathrm{Prod}(Z, \mathrm{Set}(tree, card = 6))),$

$cc1 = \mathrm{Prod}(Z, \mathrm{Union}(\mathrm{Prod}(An, tree1), tree2), \mathrm{Set}(tree, card = 4), \mathrm{Sequence}(\%1)),$

$tree1 =$

$\mathrm{Prod}(Z, \mathrm{Union}(\mathrm{Prod}(tree1, \mathrm{Set}(tree, card = 5)), \mathrm{Prod}(De, \mathrm{Set}(tree, card = 6)), De)),$

$tree2 = \mathrm{Prod}(Z, \mathrm{Union}(tree2, \mathrm{Prod}(An, tree1)), \mathrm{Set}(tree, card = 5)), cc2 = \mathrm{Prod}(Z,$
$\mathrm{Union}(\mathrm{Prod}(tree1, \mathrm{Set}(tree, card = 4)), \mathrm{Prod}(De, \mathrm{Set}(tree, card = 5)))),$
$\mathrm{Sequence}(\%1), \mathrm{Prod}(Z, An, \mathrm{Set}(tree, card = 5)), \mathrm{Sequence}(\%1)), cc = \mathrm{Cycle}(\%1)\}$
$\%1 := \mathrm{Prod}(Z, \mathrm{Set}(tree, card = 5))$

From there we compute the generating functions.

```
> sol:=gfsolve(G6,labelled,z):
```

Here is the generating function for the functional graphs:

```
> F:=subs(sol,fg(z));
```

$$F := -\frac{120}{-120 + z\,\mathrm{RootOf}(-720\,\_Z + 720\,z + z\,\_Z^6)^5}$$

Again, the coefficients admit a closed-form which can be obtained using gfun (it could also be derived by Lagrange's inversion formula). We rather compute directly the asymptotic expansion of the number of functional graphs of size n whose nodes have an in-degree of 0 or 6 by singularity analysis. Unfortunately, Maple's series command is not yet able to deal with algebraic functions like the RootOf above. We therefore start from the equation in the system which is at the origin of the singularity:

```
> T:=subs(sol,tree(z));
```

$$T := \mathrm{RootOf}(-720\,\_Z + 720\,z + z\,\_Z^6)$$

```
> tree(z)-subs(gfeqns(G6,labelled,z),tree(z));
```

$$\mathrm{tree}(z) - \mathrm{Z}(z) - \frac{1}{720}\mathrm{Z}(z)\,\mathrm{tree}(z)^6$$

```
> eq:=subs(tree(z)=y,Z(z)=z,");
```

$$eq := y - z - \frac{1}{720} z\, y^6$$

```
> diff(eq,y);
```

$$1 - \frac{1}{120} z\, y^5$$

```
> solve({"",""},{y,z});
```

$$\{z = \frac{5}{6}\mathrm{RootOf}(\_Z^3 - 12), y = \mathrm{RootOf}(\_Z^3 - 12)\},$$
$$\{y = \mathrm{RootOf}(\_Z^3 + 12), z = \frac{5}{6}\mathrm{RootOf}(\_Z^3 + 12)\}$$

It follows that the dominant singularity is at $z = \frac{5\,12^{\left(\frac{1}{3}\right)}}{6}$, where $y = 12^{\left(\frac{1}{3}\right)}$. The singular expansion of $\mathrm{y}(z)$ can be computed at that point by power series reversion (we set $\theta = \sqrt{1 - \frac{z}{\rho}}$ where ρ is the singularity, which helps series a little):

```
> sing:=5/6*12^(1/3):
> series(subs(z=sing*(1-theta^2),eq),y=12^(1/3));
```

$$(12^{1/3} - 12^{1/3}(1 - \theta^2)) + \theta^2 (y - 12^{1/3}) - \frac{5}{24} 12^{2/3} (1 - \theta^2)(y - 12^{1/3})^2 -$$
$$\frac{5}{18} 12^{1/3} (1 - \theta^2)(y - 12^{1/3})^3 + \left(-\frac{5}{24} + \frac{5}{24}\theta^2\right)(y - 12^{1/3})^4 - \frac{1}{144} 12^{2/3} (1 - \theta^2)$$

$$(y-12^{1/3})^5+O((y-12^{1/3})^6)$$

```
> {solve(",y)};
```

$$\{12^{1/3}-\frac{1}{5}\sqrt{10}\;12^{1/3}\,\theta-\frac{1}{15}12^{1/3}\,\theta^2-\frac{71}{900}\sqrt{10}\;12^{1/3}\,\theta^3+\frac{7}{3375}12^{1/3}\,\theta^4+O(\theta^5),$$
$$12^{1/3}+\frac{1}{5}\sqrt{10}\;12^{1/3}\,\theta-\frac{1}{15}12^{1/3}\,\theta^2+\frac{71}{900}\sqrt{10}\;12^{1/3}\,\theta^3+\frac{7}{3375}12^{1/3}\,\theta^4+O(\theta^5)\}$$

```
> sery:=op(select(proc(s,u) evalb(signum(coeff(s,u,1))=-1)
  end,",theta));
```

$$sery:=$$
$$12^{1/3}-\frac{1}{5}\sqrt{10}\;12^{1/3}\,\theta-\frac{1}{15}12^{1/3}\,\theta^2-\frac{71}{900}\sqrt{10}\;12^{1/3}\,\theta^3+\frac{7}{3375}12^{1/3}\,\theta^4+O(\theta^5)$$

```
> series(subs(T=sery,z=sing*(1-theta^2),F),theta);
```

$$\frac{1}{10}\sqrt{10}\;\theta^{-1}+\frac{4}{15}+\frac{47}{600}\sqrt{10}\;\theta+\frac{616}{3375}\theta^2+O(\theta^3)$$

Hence the number of functional graphs with in-degree in {0,6} is asymptotically:

```
> as_fg:=coeff(",theta,-1)/sqrt(Pi)*n^(-1/2)*sing^(-n);
```

$$as\_fg:=\frac{1}{10}\frac{\sqrt{10}\left(\frac{5}{6}12^{1/3}\right)^{(-n)}}{\sqrt{\pi}\sqrt{n}}$$

We next consider the distance to the cycle

```
> Fpl:=subs(sol,fg1(z));
```

$$Fpl:=150\frac{z\,\%1^4\,(-5\,\%1+6\,z)}{(-30-26\,z\,\%1^5+99\,z^2\,\%1^4-126\,z^3\,\%1^3+54\,z^4\,\%1^2)\,(-120+z\,\%1^5)}$$
$$\%1:=\mathrm{RootOf}(-720\,\_Z+720\,z+z\,\_Z^6)$$

```
> series(subs(T=sery,z=sing*(1-theta^2),Fpl),theta);
```

$$\frac{1}{20}\theta^{-4}+O(\theta^{-3})$$

And the length of the cycle:

```
> Fpl2:=subs(sol,fg2(z));
```

$$Fpl2:=-1440\frac{z^2\,\%1^3\,(-\%1+z)}{(-40+7\,z\,\%1^5-18\,z^2\,\%1^4+12\,z^3\,\%1^3)\,(-120+z\,\%1^5)}$$
$$\%1:=\mathrm{RootOf}(-720\,\_Z+720\,z+z\,\_Z^6)$$

```
> series(subs(T=sery,z=sing*(1-theta^2),Fpl2),theta);
```

$$\frac{1}{20}\theta^{-4}-\frac{11}{300}\sqrt{10}\;\theta^{-3}+O(\theta^{-2})$$

Thus the distance to the cycle and the length of this cycle are both asymptotic to:

```
> asympt(coeff(",theta,-4)*n*sing^(-n)/as_fg/n,n);
```

$$\frac{1}{20}\frac{\sqrt{10}\sqrt{\pi}}{\sqrt{\frac{1}{n}}}$$

Thus the ratio between this variant of Pollard's algorithm and the original one is in both cases (number of steps to the cycle and length of the cycle):

```
> ratio:=simplify(map(asympt,[as_cost_1/",as_cost_2/"/2],n))
  ;
```

$$ratio := [\sqrt{5}, \sqrt{5}]$$

Conclusion

This worksheet shows that some algorithms whose complexity analysis does not reduce to counting the number of subcomponents in some kind of recursive combinatorial structure can sometimes be treated by combstruct using several marks. This is in particular true of many recursive searching algorithms whose complexity is related to the path length of an underlying structure.