Refined upper bounds on the size of the condensed neighbourhood of sequences
Cedric Chauve, May 2021.
Preliminary version: France Paquet-Nadeau, July 2017.

```
> restart:
  with(gfun):
  march('open',
  "C:\\Users\\cchau\\Desktop\\SFU\\PSC_2021\\CondensedNeighbourhoods
  \\algolib.mla"):
```

Variables:
k = number of symbols in the query of the edit script
s = size of alphabet (default: 2)
d = distance bound, maximum number of edit operations in edit scripts

We use the two functions below GenRec and S to evaluate S(s, k, d).
The function S will be kept symbolic while GenRec will evaluate each term;
S1 is used for S in GenRec.
For actual evaluation we assume that s has been given an integer value.

```
> GenRec:= proc(s, k, d::integer)
  local j:
  if d = 0 or (type(k, posint) and k <= d) then
    return 1
  else
    return S1(s, k-1, d)
    + (s-1) * S1(s, k-1, d-1)
    + (s-1) * add(s^j * S1(s, k-2, d-1-j), j=0..d-1)
    + (s-1)^2 * add(s^j * S1(s, k-2, d-2-j), j=0..d-2)
    + add(S1(s, k-2-j, d-1-j), j=0..d-1):
  fi:
  end proc:
```

```
> S:= proc(s, k, d::integer)
  local j:
  option remember;
  if d = 0 then
    return 1
  elif type(k, nonnegint) then
    if k <= d then
      return 1
    else
      return procname (s, k-1, d)
      +(s-1) * procname(s, k-1, d-1)
      +(s-1) * add(s^j * procname(s, k-2, d-1-j), j=0..d-1)
      +(s-1)^2 * add(s^j * procname(s, k-2, d-2-j), j=0..d-2)
      + add(procname(s, k-2-j, d-1-j), j=0..d-1):
    fi:
  else
    'S'(args):
  fi:
  end proc:
```

Transforming the recurrences of S into generating function.
We will denote by F[d] the gen. functions S_{s,d}(z) obtained from this method for a given value of d.
To obtain F[1] we compute S(s, k, 1) and input the resulting expression manually into ***rectodiffeq***
to solve the differential expression defined by the recurrence.
For the following values of d, we construct 'expression' with the recurrences of 'S'.

```
> F ≔ 'F':  # Initializing
> r ≔ 9 : # Rank  of the gen. function we compute
> F[1] :=
  factor(
    solve(
      rectodiffeq({a(n)=a(n-1)+(2*s)-1, a(0)=1}, a(n), f(z)),
      f(z)
    )
  );
```

$$F_1 := \frac{2\,z\,s - 2\,z + 1}{(z-1)^2} \qquad (1)$$

We now iterate building F[d], i.e. S_{s,d}(z), from lower values of d.
We look at the gen. functions for d from 2 to 4.

```
> d_max ≔ 4 :
> for d from 2 to d_max do: # dPrime used as an index to build the
  table of gen. functions
    # RHS of the equation of S(s, k,d) with similar terms
    # grouped together
    expression :=
    S(s, k, d)=collect(eval(subs(S1=S, GenRec(s, k, d))), S):
    # Collecting the constant terms
    const := eval(subs(S=(()-> 0), expression)):
    # Removing the constant terms from the equation to keep
    # only terms of the form 'S(s, k, d)'
    expression := expression - const;
    # Transforming 'expression' into a gen. fun.
    for dd from 1 to d do
      # dd corresponds to the distance in the terms 'S(s, k, d)'
      for j from 0 to d do
        # j is the index for terms substracted to 'k'
        # Substitute the term S(s, k-j, dd) with the equation
        # for F[dd] adjusted to the index
        expression :=
        subs(
          S(s, k-j, dd)=z^j*(F[dd]-(add (eval(S(s,l+1,dd))*z^l,
          l=0..(d-j-1)))),
          expression
        ):
      od:
    od:
    # Adding back the constant terms transformed into gen. fun.
    expression := expression +  const * (z^d / (1-z)):
    F[d] := factor(solve(expression, F[d]));
    print(simplify(F[d]));
  od:
```

$$\frac{-1 + z^3 + \left(-4\,s^2 + 4\,s - 3\right) z^2 + 2\,z}{(z-1)^3}$$

$$\left[\begin{array}{l}\frac{1}{(z-1)^4}\left(1+\left(2 s^3-s^2-3 s+3\right) z^5+\left(-4 s^3+6 s^2-2 s-3\right) z^4+\left(10 s^3-17 s^2\right.\right.\\ \qquad \left.\left.+11 s-2\right) z^3+3 z^2-3 z\right)\\ \frac{1}{(z-1)^5}\left(-1+\left(4 s^4-8 s^3+8 s^2-6 s+3\right) z^7+\left(-20 s^4+32 s^3-14 s^2+4 s-6\right) z^6\right.\\ \qquad +\left(28 s^4-56 s^3+40 s^2-14 s+8\right) z^5+\left(-28 s^4+64 s^3-58 s^2+24 s-7\right) z^4+4 z^3\\ \qquad \left.-6 z^2+4 z\right)\end{array}\right. \tag{2}$$

We now define T_{s,d}(z) parameterized by d as we will consider s as fixed and z is the formal variable of the gen. functions. We show the value of T_{s,d}(z) for d from 1 to 5.

```
> for d from 1 to d_max do:
    T[d]:=F[d]+add(s^j*z*(F[d-j]-1), j=1..d-1)+(s^d / (1 - z)):
    print(simplify(T[d]))
  od:
```

$$\frac{(s-2)z+s+1}{(z-1)^2}$$

$$\frac{-s z^4+\left(2 s^2+s+1\right) z^3+\left(-7 s^2+4 s-3\right) z^2+\left(2 s^2+2\right) z-s^2-1}{(z-1)^3}$$

$$\frac{1}{(z-1)^4}\left(\left(2 s^3-2 s^2-3 s+3\right) z^5+\left(-6 s^3+12 s^2-2 s-3\right) z^4+\left(9 s^3-22 s^2+10 s-2\right) z^3+\left(5 s^3+s+3\right) z^2+\left(-3 s^3-3\right) z+s^3+1\right)$$

$$\frac{1}{(z-1)^5}\left(\left(6 s^4-9 s^3+5 s^2-3 s+3\right) z^7+\left(-26 s^4+38 s^3-13 s^2-3 s-6\right) z^6+\left(40 s^4-72 s^3+53 s^2-8 s+8\right) z^5+\left(-37 s^4+70 s^3-70 s^2+19 s-7\right) z^4+\left(6 s^4+5 s^3+2 s^2+4 s+4\right) z^3+\left(-8 s^4-s^2-s-6\right) z^2+\left(4 s^4+4\right) z-s^4-1\right) \quad \textbf{(3)}$$

We now turn to the asymptotics analysis.
The code below takes as input an array GF of gen. fun. prameterized by d
and depending on s, a maximum value for d and an interval of values for s and compute
the asymptotics of the gen. functions S (array F).
The main tool for this computation is the function "equivalent" from the Maple library
algolib (http://algo.inria.fr/libraries/).

```
> print("s, d, asymptotics of S");
  for s from 2 to 6 do
    for d from 1 to d_max do
      S_A := equivalent(F[d], z, k);
      print(s, d,
        map(z → ifactor(numer(z)) / ifactor(denom(z)),
            coeff(convert(S_A*d!, polynom), k, d)) *k^d
      )
    od:
  od:
```

"s, d, asymptotics of S"

$2, 1, (3)\,k$

$2, 2, (3)^2\,k^2$

$2, 3, (3)^3\,k^3$

$2, 4, (3)^4\,k^4$

$3, 1, (5)\,k$

$3, 2, (5)^2\,k^2$

$3, 3, (5)^3\,k^3$

$3, 4, (5)^4\,k^4$

$4, 1, (7)\,k$

$4, 2, (7)^2\,k^2$

$4, 3, (7)^3\,k^3$

$4, 4, (7)^4\,k^4$

$5, 1, (3)^2\,k$

$5, 2, (3)^4\,k^2$

$5, 3, (3)^6\,k^3$

$5, 4, (3)^8\,k^4$

$6, 1, (11)\,k$

$6, 2, (11)^2\,k^2$

$6, 3, (11)^3\,k^3$

$6, 4, (11)^4\,k^4$ **(4)**

We repeat the same analysis with the gen. function T.

```
> print("s, d, asymptotics of T");
  for s from 2 to 6 do
    for d from 1 to d_max do
      T_A := equivalent(T[d], z, k);
      print(s, d,
        map(z → ifactor(numer(z))/ifactor(denom(z)),
            coeff(convert(T_A*d!, polynom), k, d))*k^d
      )
    od:
  od:
```

"s, d, asymptotics of T"

$$2, 1, (3)\, k$$
$$2, 2, (3)^2\, k^2$$
$$2, 3, (3)^3\, k^3$$
$$2, 4, (3)^4\, k^4$$
$$3, 1, (5)\, k$$
$$3, 2, (5)^2\, k^2$$
$$3, 3, (5)^3\, k^3$$
$$3, 4, (5)^4\, k^4$$
$$4, 1, (7)\, k$$
$$4, 2, (7)^2\, k^2$$
$$4, 3, (7)^3\, k^3$$
$$4, 4, (7)^4\, k^4$$
$$5, 1, (3)^2\, k$$
$$5, 2, (3)^4\, k^2$$
$$5, 3, (3)^6\, k^3$$
$$5, 4, (3)^8\, k^4$$
$$6, 1, (11)\, k$$
$$6, 2, (11)^2\, k^2$$
$$6, 3, (11)^3\, k^3$$
$$6, 4, (11)^4\, k^4 \quad \textbf{(5)}$$