

# Conjecture de Goldbach, langage, réécriture

Denise Vella-Chemla

9/2/14

## 1 Introduction

On souhaite trouver une démonstration de la conjecture de Goldbach, qui stipule que tout nombre pair supérieur à 2 est la somme de deux nombres premiers<sup>1</sup>.

On se propose dans ce but d'utiliser une modélisation qui associe à chaque nombre pair  $n$  un mot d'un langage à 4 lettres qui code la primalité des nombres impairs  $x$  (compris entre 3 et  $n/2$ ) et de leur complémentaire.

On identifiera le processus permettant de passer du mot d'un nombre pair  $n$  au mot du nombre pair suivant  $n + 2$ .

On caractérisera l'existence d'un décomposant de Goldbach d'un nombre pair par une simple condition que vérifie son mot.

On essaiera de trouver une contrainte invariante respectée par les mots des nombres pairs successifs qui assurera que l'existence d'un décomposant de Goldbach est toujours conservée.

## 2 Mots d'un nombre pair

On choisit de représenter le fait qu'un entier est premier par le booléen 0 et le fait qu'il est composé par le booléen 1.

On appelle  $sym(m)$  la fonction qui associe à un mot  $m$  son symétrique, i.e. le mot contenant les lettres de  $m$  depuis la dernière jusqu'à la première.

Appelons *milieu* le plus grand nombre entier impair inférieur ou égal à  $n/2$ .

---

1. Dans l'égalité  $n = p + q$  avec  $n$  pair supérieur à 2,  $p$  et  $q$  premiers, on appellera  $p$  et  $q$  décomposants de Goldbach de  $n$  ou sommants.

A chaque nombre pair  $n$  sont associés deux mots booléens  $m_1$  et  $m_2$  définis de la façon suivante :

- les lettres de  $m_1$  sont les caractères de primalité des nombres impairs compris entre 3 et *milieu* inclus ;
- les lettres de  $m_2$  sont les caractères de primalité des nombres impairs compris entre  $n - 3$  et *milieu* inclus.

Les mots  $m_1$  et  $m_2$  associés au nombre pair  $n$  sont de longueur  $\lfloor \frac{n/2 - 1}{2} \rfloor$ . La longueur des mots augmente donc de 1 à chaque double d'impair, i.e. une fois sur deux.

Le mot booléen  $m$  du nombre pair  $n$  est la concaténation des deux mots suivants :

- $m_1$  ;
- $sym(m_2)$ , le symétrique de  $m_2$ .

*Note* : on a pris pour habitude de fournir le mot  $m_2$  en première ligne et le mot  $m_1$  en deuxième ligne (3 en bas à gauche,  $n - 3$  en haut à gauche). On constate que pour les doubles d'impairs, la lettre codant le caractère de primalité de l'entier *milieu* est doublée.

On associe d'autre part à  $n$  un mot d'un langage à 4 lettres  $m_{abcd}$ , dont chaque lettre code les colonnes de lettres des mots  $m_2$  et  $m_1$ .

La lettre  $a$  code la colonne  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ . La lettre  $b$  code la colonne  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . La lettre  $c$  code la colonne  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ . La lettre  $d$  code la colonne  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

**Exemples** : Ci-dessous les mots  $m_1$ ,  $m_2$ ,  $m$  et  $m_{abcd}$  des nombres 40, 42 et 44.

40	37	35	33	31	29	27	25	23	21										
$m_2$	0	1	1	0	0	1	1	0	1										
$m_1$	0	0	0	1	0	0	1	0	0										
	3	5	7	9	11	13	15	17	19										
$m$	0	0	0	1	0	0	1	0	0		1	0	1	1	0	0	1	1	0
$m_{abcd}$	$a$	$c$	$c$	$b$	$a$	$c$	$d$	$a$	$c$										

42	39	37	35	33	31	29	27	25	23	21											
$m_2$	1	0	1	1	0	0	1	1	0	1											
$m_1$	0	0	0	1	0	0	1	0	0	1											
	3	5	7	9	11	13	15	17	19	21											
$m$	0	0	0	1	0	0	1	0	0	1		1	0	1	1	0	0	1	1	0	1
$m_{abcd}$	$c$	$a$	$c$	$d$	$a$	$a$	$d$	$c$	$a$	$d$											

44	41	39	37	35	33	31	29	27	25	23											
$m_2$	0	1	0	1	1	0	0	1	1	0											
$m_1$	0	0	0	1	0	0	1	0	0	1											
	3	5	7	9	11	13	15	17	19	21											
$m$	0	0	0	1	0	0	1	0	0	1		0	1	1	0	0	1	1	0	1	0
$m_{abcd}$	$a$	$c$	$a$	$d$	$c$	$a$	$b$	$c$	$c$	$b$											

### 3 Identifier ce que fait le processus

Reprenons les mots  $m_1$ ,  $m_2$  et  $m$  des nombres pairs 24 à 34.

24	$m_2$	1	0	0	1	0															
	$m_1$	0	0	0	1	0															
	$m$	0	0	0	1	0		0	1	0	0	1									
26	$m_2$	0	1	0	0	1	0														
	$m_1$	0	0	0	1	0	0														
	$m$	0	0	0	1	0	0		0	1	0	0	1	0							
28	$m_2$	1	0	1	0	0	1														
	$m_1$	0	0	0	1	0	0														
	$m$	0	0	0	1	0	0		1	0	0	1	0	1							
30	$m_2$	1	1	0	1	0	0	1													
	$m_1$	0	0	0	1	0	0	1													
	$m$	0	0	0	1	0	0	1		1	0	0	1	0	1	1					
32	$m_2$	0	1	1	0	1	0	0													
	$m_1$	0	0	0	1	0	0	1													
	$m$	0	0	0	1	0	0	1		0	0	1	0	1	1	0					
34	$m_2$	0	0	1	1	0	1	0	0												
	$m_1$	0	0	0	1	0	0	1	0												
	$m$	0	0	0	1	0	0	1	0		0	0	1	0	1	1	0	0			

On voit que si au nombre pair  $n$  est associé un mot booléen de longueur  $2i$ , le processus qui permet d'obtenir le mot booléen associé au nombre pair  $n + 2$  effectue plusieurs actions différentes :

- *travail sur la lettre à la position  $i + 1$*  (on a coloré cette lettre en bleu dans le tableau ci-dessus) : dans le cas où  $n$  est un double d'impair, le mot de  $n + 2$  est obtenu en enlevant du mot de  $n$  la lettre à la position  $i + 1$  ; dans le cas où  $n$  est un double de pair, le mot de  $n + 2$  est obtenu en dupliquant cette lettre à la position  $i + 1$  ;
- *concaténation en fin du mot* : dans tous les cas, est concaténée à la fin du mot booléen ainsi obtenu la lettre qui caractérise la primalité du nombre entier  $2n - 3$  pour obtenir le mot de  $n + 2$ . *Remarque* : la concaténation est une opération non-commutative. Par exemple,  $1(110) = 1110$  alors que  $(110)1 = 1101$ .

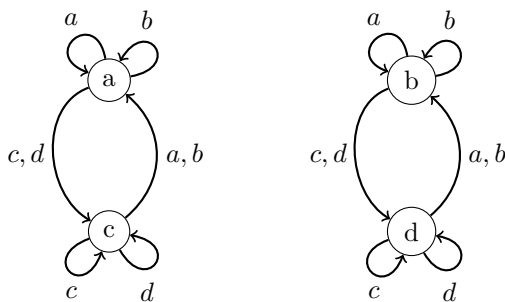
Si on considère maintenant la représentation par les mots du langage à 4 lettres, celui du nombre pair  $n + 2$  s'obtient de la façon suivante à partir de celui de  $n$  :

- la première lettre du mot de  $n + 2$  est  $a$  si  $n - 3$  est premier et  $c$  sinon (cette première lettre est la seule qui introduit de l'indéterminisme car elle n'appartient pas au mot du langage à 4 lettres de  $n$  ou ne se déduit pas des lettres de ce dernier) ;
- les lettres suivantes du mot de  $n + 2$  sont obtenues par réécriture du mot de  $n$  selon les règles ci-dessous :

$aa \rightarrow a$   
 $ab \rightarrow b$   
 $ac \rightarrow a$   
 $ad \rightarrow b$   
 $ba \rightarrow a$   
 $bb \rightarrow b$   
 $bc \rightarrow a$   
 $bd \rightarrow b$   
 $ca \rightarrow c$   
 $cb \rightarrow d$   
 $cc \rightarrow c$   
 $cd \rightarrow d$   
 $da \rightarrow c$   
 $db \rightarrow d$   
 $dc \rightarrow c$   
 $dd \rightarrow d$

On note que 4 règles de réécriture ( $aa \rightarrow a$ ,  $ac \rightarrow a$ ,  $ba \rightarrow a$ ,  $bc \rightarrow a$ ) assurent d'obtenir une lettre  $a$  au moins dans le mot de  $n + 2$ .

On peut représenter ces règles de réécriture par les deux petits automates déterministes suivants (l'opérateur est à gauche) :



- enfin, la concaténation d'une lettre en fin de mot, dans le cas où  $n$  est un double de pair obéit à la règle suivante :
  - si  $n$  a  $a$  ou  $b$  comme dernière lettre, après avoir obtenu le mot de  $n + 2$  en appliquant les règles de réécriture, on lui concatène la lettre  $a$  ;
  - si  $n$  a  $c$  ou  $d$  comme dernière lettre, après avoir obtenu le mot de  $n + 2$  en appliquant les règles de réécriture, on lui concatène la lettre  $d$ .

## 4 Loi de composition de Ritz-Rydberg

On teste ici sur deux exemples la loi de composition de Ritz-Rydberg, qui a pour conséquence que la composition des règles de réécriture  $(\alpha, \beta)$  et  $(\beta, \gamma)$  a le même effet que la règle de réécriture  $(\alpha, \gamma)$ .

$ab/bc \rightarrow bba \rightarrow ba \rightarrow a$  permet d'obtenir le même résultat que  $ac \rightarrow a$ .

$cd/da \rightarrow ddc \rightarrow dc \rightarrow c$  permet d'obtenir le même résultat que  $ca \rightarrow c$ .

En annexe 2, sont fournies les 64 règles de composition qui vérifient le principe de Ritz-Rydberg.

## 5 Caractériser l'existence d'une décomposition de Goldbach dans le mot d'un nombre pair

Il faut maintenant être capable de caractériser par une condition sur le mot  $m$  la présence à une même position dans les mots  $m_1$  et  $m_2$  d'une lettre 0.

Rappelons quelques éléments de logique booléenne.

La conjonction logique est définie par :

$$1 \wedge 0 = 0 \wedge 1 = 0 \wedge 0 = 0 \text{ et } 1 \wedge 1 = 1.$$

La négation logique est définie par :

$$\neg 0 = 1 \text{ et } \neg 1 = 0.$$

Si l'on appelle  $l(m, i)$  la lettre à la position  $i$  dans le mot  $m$ , alors l'existence d'une décomposition de Goldbach est équivalente à la condition :

$$\left[ \sum_{1 \leq i \leq \lfloor \frac{n/2-1}{2} \rfloor, i+j=\lfloor \frac{n}{2} \rfloor} \neg l(m, i) \wedge \neg l(m, j) \right] = 1$$

En utilisant la représentation par les mots du langage à 4 lettres, l'existence d'une décomposition de Goldbach est simplement la présence d'une lettre  $a$  dans le mot du nombre pair considéré.

Le double d'un nombre impair dont le mot  $m_{abcd}$  se termine par une lettre  $a$  est un double de nombre premier, qui vérifie donc trivialement la conjecture de Goldbach (ex : 46 dont le mot  $m_{abcd}$  est  $aacbcbacda$  se terminant par une lettre  $a$  est le double de 23, premier).

Le double d'un nombre pair dont le mot  $m_{abcd}$  se termine par une lettre  $a$  est le double d'un "père de jumeau" (ex : 36 dont le mot  $m_{abcd}$  est  $acabca$  se terminant par une lettre  $a$  est le double de 18, un père de jumeau, i.e. un nombre pair compris entre deux nombres premiers, en l'occurrence 17 et 19).

## 6 Invariant

Rappelons un fait qui peut-être être utile : tous les mots  $m_{abcd}$  des nombres pairs que nous considérerons ne pourront jamais contenir une suite de 3 lettres  $a$  consécutives : cela provient du fait que 3, 5 et 7 sont les seuls trois nombres premiers consécutifs puisque toute suite de trois impairs consécutifs contient un nombre divisible par 3.

On constate dans l'annexe que les mots contiennent des "lettres alignées" selon des verticales ou des diagonales descendantes, que ce soit des  $b$  ou  $d$  d'une part, ou des  $a$  ou  $c$  d'autre part.

Ces lignes sont vite identifiées comme correspondant aux décompositions successives faisant soit intervenir le même premier sommant, soit intervenir le même deuxième sommant. La quatrième ligne verticale de lettres par exemple, qui commence par les lettres  $dbb dbdb...$  correspond aux décompositions  $9 + 9, 9 + 11, 9 + 13, \dots$ . La sixième diagonale descendante, à partir de la première lettre du mot de 18, et qui contient les lettres  $cccdccd$  correspond aux décompositions  $3 + 15, 5 + 15, 7 + 15, 9 + 15, \dots$

On peut donc de manière imagée, considérer que le mot  $m_{abcd}$  d'un nombre pair est une sorte de sandwich multi-couches, qui contient une première tranche de lettres  $a$  ou  $c$  en début de mot, suivie d'un certain nombre de tranches alternées, les unes ne contenant que des lettres  $b$  ou  $d$  et les autres ne contenant que des lettres  $a$  ou  $c$ , et que les tranches voient les positions de leur première et dernière lettre fixées une fois pour toutes, même si leur composition varie au fur et à mesure du processus.

Supposons qu'un mot  $n + 2$  n'admette pas de décomposition de Goldbach. Ce mot ne doit contenir que des lettres  $c, b$  ou  $d$ . On note ce mot de la façon suivante, par abus de notation :  $(c^*(b \vee d)^*)^*$ .

Essayons de trouver d'où pourrait provenir la contradiction si on prend comme hypothèse l'existence d'un tel mot.

Pour cela, essayons d'imaginer la composition du mot  $m_{abcd}$  associé au nombre pair précédent qui est  $n$ . Il pourrait contenir des lettres  $a$  mais elles devraient forcément être en "fin des tranches" ( $c \vee a$ ) puisque sinon, toute occurrence du doublon de lettres  $ac$  entraînerait la présence d'une lettre  $a$  dans le mot du nombre pair  $n + 2$ , ce qui serait contraire à l'hypothèse. Un raisonnement similaire oblige les lettres  $d$  à être à la fin des tranches ( $b \vee d$ ). Le mot du nombre pair  $n$  serait donc obligatoirement de la forme  $(c^*a(b \vee d)^*d)^*$ . Il contiendrait des lettres  $a$ . On n'arrive pas à aboutir à une contradiction pour l'instant.

Pour regarder autrement le processus, voyons comment les règles de réécriture se combinent entre elles, dans un tableau de concaténation : si on concatène  $xy$  avec  $zt$ , on va s'intéresser à la lettre obtenue par le "jointure" des deux doublons de lettres, c'est à dire qu'on notera dans le tableau le résultat de la règle de réécriture qui a les lettres  $yz$  en partie gauche :

	<i>ab</i>	<i>ad</i>	<i>bb</i>	<i>bd</i>	<i>ca</i>	<i>cb</i>	<i>cc</i>	<i>cd</i>	<i>da</i>	<i>db</i>	<i>dc</i>	<i>dd</i>
<i>ab</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>ad</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>bb</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>bd</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>ca</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>cb</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>cc</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>cd</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>da</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>db</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
<i>dc</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>
<i>dd</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>

On n'arrive pas à avancer dans cette voie-là non plus.

Si on calcule la proportion de lettres *a* (de décompositions de Goldbach donc) dans les mots  $m_{abcd}$  de  $n$  valant certaines puissances de 10, on trouve qu'elle est égale à  $5402/249999 = 0.0216$  pour  $n = 10^6$ ,  $38807/2499999 = 0.0155$  pour  $n = 10^7$  et  $291400/249999 = 0.0116$  pour  $n = 10^8$ . Elle semble ne même pas être divisée par 2 lorsqu'on multiplie  $n$  par 10 mais cela ne prouve rien.

Si enfin on compte simplement le nombre d'occurrences de chaque lettre en partie gauche ou droite des règles de réécriture, les lettres sont bien sûr présentes 8 fois en tout dans les 16 parties gauches, et équitablement 4 fois chacune en partie droite des règles. Mais les règles sont comme "entremêlées" ce qui empêche de trouver une fonction qui permettrait d'identifier certaines d'entre elles (les lettres *a* et *b* par exemple, se comportent de la même manière en tant que préfixes des parties gauches, mais totalement différemment en tant que suffixes).

## Annexe 1 : mots du langage à 4 lettres associés aux nombres pairs de 6 à 100

6 : a  
 8 : a  
 10 : a a  
 12 : c a  
 14 : a c a  
 16 : a a c  
 18 : c a a d  
 20 : a c a b  
 22 : a a c b a  
 24 : c a a d a  
 26 : a c a b c a  
 28 : c a c b a c  
 30 : c c a d a a d  
 32 : a c c b c a b  
 34 : a a c d a c b a  
 36 : c a a d c a d a  
 38 : c c a b c c b c a  
 40 : a c c b a c d a c  
 42 : c a c d a a d c a d  
 44 : a c a d c a b c c b  
 46 : a a c b c c b a c d a  
 48 : c a a d a c d a a d c  
 50 : a c a b c a d c a b c d  
 52 : c a c b a c b c c b a d  
 54 : c c a d a a d a c d a b d  
 56 : a c c b c a b c a d c b b  
 58 : c a c d a c b a c b c d b a  
 60 : c c a d c a d a a d a d d a  
 62 : a c c b c c b c a b c b d c a  
 64 : a a c d a c d a c b a d b c c  
 66 : c a a d c a d c a d a b d a c d  
 68 : c c a b c c b c c b c b b c a d  
 70 : a c c b a c d a c d a d b a c b d  
 72 : c a c d a a d c a d c b d a a d b  
 74 : a c a d c a b c c b c d b c a b d a  
 76 : a a c b c c b a c d a d d a c b b c  
 78 : c a a d a c d a a d c b d c a d b a d  
 80 : c c a b c a d c a b c d b c c b d a b  
 82 : a c c b a c b c c b a d d a c d b c b a  
 84 : c a c d a a d a c d a b d c a d d a d a  
 86 : a c a d c a b c a d c b b c c b d c b c a  
 88 : c a c b c c b a c b c d b a c d b c d a c  
 90 : c c a d a c d a a d a d d a a d d a d c a d  
 92 : a c c b c a d c a b c b d c a b d c b c c b  
 94 : c a c d a c b c c b a d b c c b b c d a c d a  
 96 : c c a d c a d a c d a b d a c d b a d c a d c  
 98 : c c c b c c b c a d c b b c a d d a b c c b c d  
 100 : a c c d a c d a c c c c d b a c b d c b a c d a d





$cd/da \rightarrow ddc \rightarrow dc \rightarrow c$  permet d'obtenir le même résultat que  $ca \rightarrow c$ .  
 $cd/db \rightarrow ddd \rightarrow dd \rightarrow d$  permet d'obtenir le même résultat que  $cb \rightarrow d$ .  
 $cd/dc \rightarrow ddc \rightarrow dc \rightarrow c$  permet d'obtenir le même résultat que  $cc \rightarrow c$ .  
 $cd/dd \rightarrow ddd \rightarrow dd \rightarrow d$  permet d'obtenir le même résultat que  $cd \rightarrow d$ .

$da/aa \rightarrow caa \rightarrow ca \rightarrow c$  permet d'obtenir le même résultat que  $da \rightarrow c$ .  
 $da/ab \rightarrow cab \rightarrow cb \rightarrow d$  permet d'obtenir le même résultat que  $db \rightarrow d$ .  
 $da/ac \rightarrow caa \rightarrow ca \rightarrow c$  permet d'obtenir le même résultat que  $dc \rightarrow c$ .  
 $da/ad \rightarrow cab \rightarrow cb \rightarrow d$  permet d'obtenir le même résultat que  $dd \rightarrow d$ .  
 $db/aa \rightarrow dba \rightarrow da \rightarrow c$  permet d'obtenir le même résultat que  $da \rightarrow c$ .  
 $db/ab \rightarrow dbb \rightarrow db \rightarrow d$  permet d'obtenir le même résultat que  $db \rightarrow d$ .  
 $db/ac \rightarrow dba \rightarrow da \rightarrow c$  permet d'obtenir le même résultat que  $dc \rightarrow c$ .  
 $db/ad \rightarrow dbb \rightarrow db \rightarrow d$  permet d'obtenir le même résultat que  $dd \rightarrow d$ .  
 $dc/aa \rightarrow ccc \rightarrow cc \rightarrow c$  permet d'obtenir le même résultat que  $da \rightarrow c$ .  
 $dc/ab \rightarrow ccd \rightarrow cd \rightarrow d$  permet d'obtenir le même résultat que  $db \rightarrow d$ .  
 $dc/ac \rightarrow ccc \rightarrow cc \rightarrow c$  permet d'obtenir le même résultat que  $dc \rightarrow c$ .  
 $dc/ad \rightarrow ccd \rightarrow cd \rightarrow d$  permet d'obtenir le même résultat que  $dd \rightarrow d$ .  
 $dd/aa \rightarrow ddc \rightarrow dc \rightarrow c$  permet d'obtenir le même résultat que  $da \rightarrow c$ .  
 $dd/ab \rightarrow ddd \rightarrow dd \rightarrow d$  permet d'obtenir le même résultat que  $db \rightarrow d$ .  
 $dd/ac \rightarrow ddc \rightarrow dc \rightarrow c$  permet d'obtenir le même résultat que  $dc \rightarrow c$ .  
 $dd/ad \rightarrow ddd \rightarrow dd \rightarrow d$  permet d'obtenir le même résultat que  $dd \rightarrow d$ .