HAL open science

Séparateurs Algébriques dans la Bio-systémique

Annick Valibouze, Ines Abdeljaoued, Alia Ben Kahla

To cite this version:

Annick Valibouze, Ines Abdeljaoued, Alia Ben Kahla. Séparateurs Algébriques dans la Bio-systémique. 2010. hal-00484337

HAL Id: hal-00484337

https://hal.science/hal-00484337v1

Preprint submitted on 5 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SÉPARATEURS ALGÉBRIQUES DANS LA BIO-SYSTÉMIQUE

ANNICK VALIBOUZE, INES ABDELJAOUED ET ALIA BEN KAHLA

Résumé. La modélisation des systèmes biologiques est indissociable de la compréhension des mécanismes de propagation des pathogènes et de la défense de l'hôte en biologie moléculaire. La modélisation par des outils algébriques s'est appuyée jusque là sur les bases de Gröbner. Nous proposons une méthode utilisant les séparateurs algébriques qui ne sont rien d'autre que des polynômes particuliers relevant de la théorie de Galois effective. Les séparateurs algébriques sont déterminés grâce aux modules fondamentaux (générateurs de l'idéal des relations) soit directement par un produit particulier de ces modules (à l'aide des fonctions symétriques élémentaires en ces modules), soit par combinaisons linéaires des modules fondamentaux. Ces deux approches permettent d'éviter les lourds calculs des bases de Gröbner.

Abstract. The modeling of biological systems is integral to understanding the mechanisms of spread of pathogens and host defense in molecular biology. The algebraic modeling tools hitherto relied are related to Gröbner bases. We propose two methods using algebraic separators which are nothing other than special polynomials long studied in effective Galois theory. Algebraic separators are directly determined through generators of the ideal of relations (by a particular product of basic modules using the elementary symmetric functions) or by linear combinations of basic modules. These two approaches can avoid heavy calculations of Gröbner bases.

1. Introduction

Ci-dessous un échantillon de données d'expression provenant d'un réseau métabolique (voir la signification de ces données dans [5] par exemple). Il s'agit des concentrations de 3 protéines (codés dans la première ligne du tableau 1) :

Table 1. Evolution de la concentration de 3 protéines dans le temps

Code protéine	**6348**	**5606**	**5601**
concentration à 0h	19204,3	7258,53	1456,75
concentration à 3h	30590,7	13993,1	1254,47
concentration à 6h	21466,9	9777,97	1407,16
concentration à 12h	13859	7897,34	1663,16
concentration à 24h	6506,74	6250,73	1841,73

Chacune des $n = 3$ colonnes représente une protéine i et les éléments des lignes 2 à 6 représentent la concentration x_j de la protéine j respectivement aux $m = 5$ temps 0h, 3h, 6h, 12h et 24h.

Le premier traitement des données passe par leur discrétisation ; la méthode proposée par E. Dimitrova dans [4] se base sur la théorie des graphes et s'applique à des données plus compactes [1] que celles du tableau 1. Afin de pouvoir l'appliquer, nous travaillons donc sur les données du tableau 2 :

TABLE 2. Evolution de la concentration de $n = 3$ protéines dans le temps ($m = 5$)

Code protéine	**6348**	**5606**	**5601**
0h	1.9	0.7	1.5
3h	3.1	1.4	1.3
6h	2.1	1.0	1.4
12h	1.4	0.8	1.7
24h	0.7	0.6	1.8

A partir du tableau 2, la discrétisation d'E. Dimitrova consiste à choisir un entier premier p, appelé *degré de discrétisation*, et à remplacer chaque x_i par une valeur correspondante dans $\mathbb{Z}/p\mathbb{Z}$ que nous continuerons à appeler x_i.

Exemple 1.1. Il s'agit de travailler dans le corps fini $\mathbb{Z}/p\mathbb{Z}$ (ici $p = 5$). A chaque instant $t \in [1, m]$, le vecteur $s_t = (x_1, x_2, \ldots, x_n)$ est appelé l'*état du système à l'instant* t. Nous disposons d'une suite finie $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m$ appelée *trajectoire discrète* (de longueur m). Dans notre exemple, la trajectoire est $V = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_4, \boldsymbol{s}_5\}$ où $\boldsymbol{s}_1 = (2, 1, 2)$, $\boldsymbol{s}_2 = (4, 4, 0)$, $\boldsymbol{s}_3 = (3, 3, 1)$, $\boldsymbol{s}_4 = (1, 2, 3)$ et $\boldsymbol{s}_5 = (0, 0, 4)$ correspondant respectivement aux $m = 5$ lignes du tableau 3.

TABLE 3. Exemple de discrétisation avec $n = 3$, $m = 5$ et $p = 5$

Code protéine	1 **6348**	2 **5606**	3 **5601**	
0h	2	1	2	$\leftarrow \boldsymbol{s}_1$
3h	4	4	0	$\leftarrow \boldsymbol{s}_2$
6h	3	3	1	$\leftarrow \boldsymbol{s}_3$
12h	1	2	3	$\leftarrow \boldsymbol{s}_4$
24h	0	0	4	$\leftarrow \boldsymbol{s}_5$
	$\uparrow$ t_1	$\uparrow$ t_2	$\uparrow$ t_3	

Nous cherchons la fonction $f = (f_1, \ldots, f_n)$ telle que $f(\boldsymbol{s}_i) = \boldsymbol{s}_{i+1}$ pour $i = 1, \ldots, m-1$, appelée *Système dynamique polynomial*. D'après les travaux de R. Laubenbacher et B. Stigler (voir [**?**] dans [8]) et suite au calcul d'une base de Gröbner sur la variété affine définie par la trajectoire (l'algorithme adapte celui de Buchberger-Möller qui est implémenté dans les systèmes de Calcul Formel Cocoa ou Macaulay2), nous obtenons les séparateurs (voir définition 1) respectifs de $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, $\boldsymbol{s}_3$ et $\boldsymbol{s}_4$:

1. Division par 1000 pour les deux premières protéines et par 100 pour la troisième.

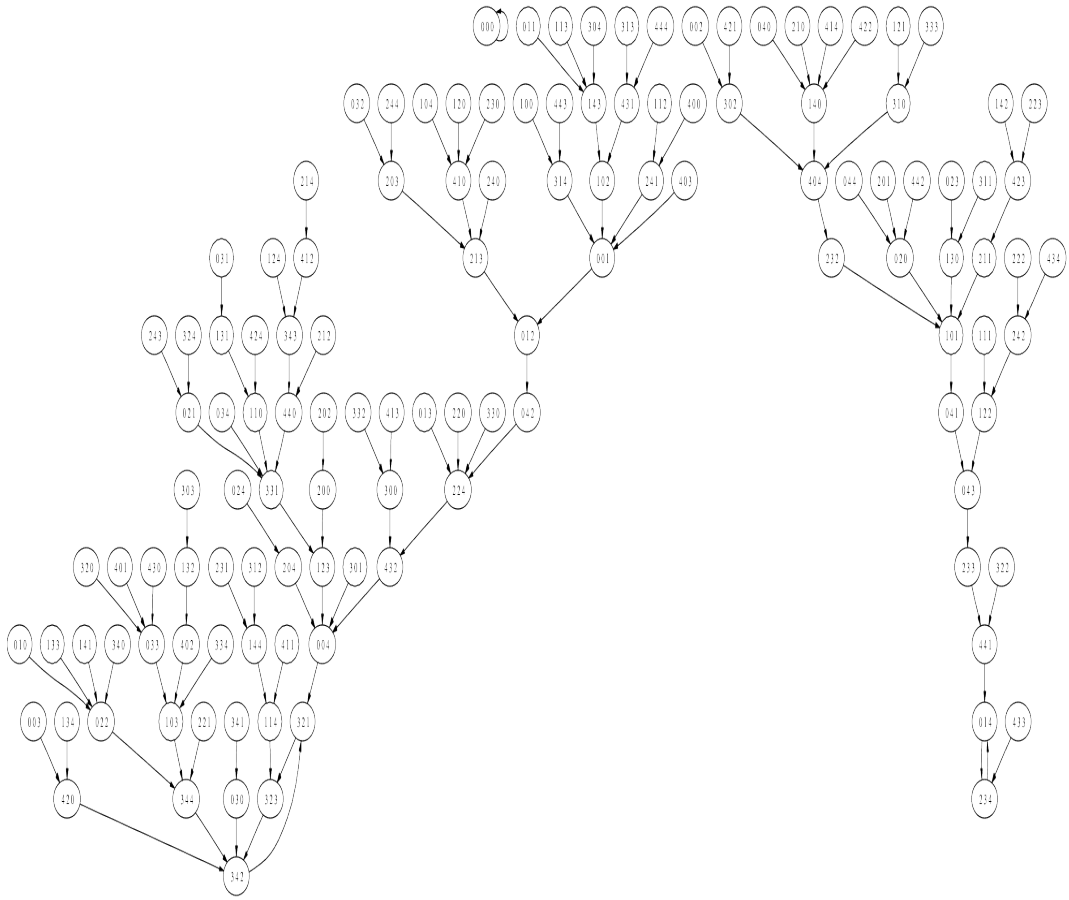

FIGURE 1. Graphe des états de transition ($n = 3$ et $p = 5$) associé à (1), (2) et (3)

$$
\begin{aligned}
r_1(x_1, x_2, x_3) &= -2x_1x_3 - 2x_2x_3 + 2x_3^2 \\
r_2(x_1, x_2, x_3) &= x_2^2 - 2x_2x_3 + 2x_3^2 \\
r_3(x_1, x_2, x_3) &= -2x_1 + 2x_2 + x_3 \\
r_4(x_1, x_2, x_3) &= -x_2x_3 - 2x_3^2
\end{aligned}
$$

Partant des concentrations respectives $t_1 = (4, 3, 1, 0)$, $t_2 = (4, 3, 2, 0)$ et $t_3 = (0, 1, 3, 4)$ des protéines 1, 2 et 3 à partir du temps 3h (voir tableau 3), nous obtenons alors :

$$
\begin{aligned}
&(1) \quad f_1(x_1, x_2, x_3) = 4r_1 + 3r_2 + r_3 + 0r_4 &&= 3x_2^2 + 2x_1x_3 + x_2x_3 - x_3^2 + 3x_1 + 2x_2 + x_3 \\
&(2) \quad f_2(x_1, x_2, x_3) = 4r_1 + 3r_2 + 2r_3 + 0r_4 &&= 3x_2^2 + 2x_1x_3 + x_2x_3 - x_3^2 + x_1 - x_2 + 2x_3 \\
&(3) \quad f_3(x_1, x_2, x_3) = 0r_1 + 1r_2 + 3r_3 + 4r_4 &&= x_2^2 - x_2x_3 - x_3^2 - x_1 + x_2 + 3x_3
\end{aligned}
$$

vérifiant $f_j(\boldsymbol{s}_i) = t_{j,i}$ pour $j = 1, 2, 3$ et $i = 1, 2, 3, 4$. A partir de la fonction $f = (f_1, f_2, f_3)$ sont déduites toutes les trajectoires, représentées par le graphe des états de transition de la figure 1.

Dans la suite de cet article, nous détaillons une approche permettant de calculer les séparateurs r_i. C'est une méthode classique partant de l'interpolation de Lagrange (voir [7]). Dans [3], les auteurs avaient évoqué l'interpolation de Lagrange sans que soit poussé l'analyse à l'élaboration d'une méthode de calcul des séparateurs algébriques : ils avaient en revanche adapté les bases de Gröbner au calcul de tous les systèmes dynamiques polynomiaux et en particulier au calcul de séparateurs algébriques. Nous présentons dans la section 2 une méthode se basant sur des outils propres à la théorie de Galois comme les modules fondamentaux ou les fonctions symétriques élémentaires (dans le cas particulier des systèmes dynamiques booléens). La section 3 présente une méthode naïve de calcul de séparateurs affines. La dernière section détaille le calcul d'un système dynamique polynomial. L'objectif étant de proposer une alternative efficace des bases de Gröbner pour une modélisation de la bio-systémique.

2. Séparateurs Algébriques

Soit k un corps fini. Un système dynamique polynomial (SDP) de dimension n est une fonction $f = (f_1, \ldots, f_n) : k^n \to k^n$ dont les composantes sont des fonctions $f_i : k^n \to k$. Remarquons que le corps k est un corps fini de cardinal $p : k = \mathbb{Z}/p\mathbb{Z}$. Ainsi, en ingénierie inversée, les polynômes f_i sont des éléments de l'anneau quotient $R = k[x_1, \ldots, x_n]/ < x_1^p - x_1, \ldots, x_n^p - x_n >$. Ce sont donc des polynômes en n variables à coefficients dans k de degré au plus égal à $p - 1$.

Soit $\mathcal{E} = k^n$ où $k = \mathbb{Z}/p\mathbb{Z}$, l'ensemble des p^n états (dans l'exemple $card(\mathcal{E}) = 5^3 = 125$). Soit $\boldsymbol{s} = (s_1, \ldots, s_n) \in \mathcal{E}$. L'idéal maximal $\mathfrak{M}_{\boldsymbol{s}}$ de $k[x_1, \ldots, x_n]$ des $\boldsymbol{s}$-relations est engendré par $\boldsymbol{p} = \{p_1, \ldots, p_n\}$:

$$p_1 = x_1 - s_1, \ldots, p_n = x_n - s_n$$

appelés par N. Tchebotarev les *modules fondamentaux* de $\boldsymbol{s}$ ([11]) ; ils satisfont pour tout $\boldsymbol{t} \in \mathcal{E}$:

$$\forall j \in [1, n] \quad p_j(\boldsymbol{t}) = 0 \quad \Leftrightarrow \quad \boldsymbol{t} = \boldsymbol{s} \quad ; \tag{4}$$

ce qui signifie que la variété de $\mathfrak{M}_{\boldsymbol{s}}$ est réduite au seul élément $\boldsymbol{s}$. C'est un cas particulier de la théorie de Galois dans lequel le groupe de Galois sur k du polynôme $(x - s_1) \cdots (x - s_n)$ est le groupe identité (voir [14]).

Définition 1. Soient $V \subset \mathcal{E}$ représentant une trajectoire discrète de longueur finie et $\boldsymbol{s} \in V$. Un polynôme $r(\boldsymbol{x})$ appartenant à $k[x_1, \ldots, x_n]$ qui vaut 1 en $\boldsymbol{x} = \boldsymbol{s}$ et 0 en les autres éléments de V est appelé un (polynôme) *séparateur* de $\boldsymbol{s}$ dans V.

Nous avons ce premier théorème non satisfaisant du point de vue de la minimalité mais qui nous éclaire sur le calcul des séparateurs algébriques ainsi que sur le second théorème qui donne un séparateur minimal [2].

Définition 2. Soit J l'ensemble des indices j pour lesquels tous les éléments de $V = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ possèdent la même j-ième coordonnée :

$$J := \{j \in \{1, \ldots, n\} \mid \forall l \in [1, m] \quad s_{l,j} = s_{1,j}\} \quad ;$$

2. Nous reviendrons plus loin sur la signification de cette minimalité.

sur les coordonnées indicées par j aucune séparation ne sera possible. Fixons $S := \{1, \ldots, n\} \backslash J$, le sous-ensemble minimal $\{1, \ldots, n\}$ séparant les éléments V ; nous l'appellerons l'*ensemble séparateur* de V.

Notons que l'ensemble J exclu les produits de gènes dont la concentration est constante dans le temps. Nous garderons pour nos calculs les produits de gènes qui varient dans le temps : ce sont les éléments de l'ensemble séparateur S. Pour $V = \mathcal{E}$ cet ensemble est $S = \{1, \ldots, n\}$.

Théorème 1. *Soit le polynôme*

$$g(\boldsymbol{x}) = \prod_{j \in S} \prod_{l \in E} (p_j(\boldsymbol{x}) - l)$$

où S est l'ensemble séparateur de V et E l'ensemble des valeurs non nulles prises par les points de V en les polynômes engendrant l'idéal des $\boldsymbol{s}$-relations (hormis ceux inutiles à la séparation) :

$$E = \{p_j(\boldsymbol{t}) \mid j \in S \; ; \; p_j(\boldsymbol{t}) \neq 0 \; ; \; \boldsymbol{t} \in V\} \subset \{1, \ldots, p-1\} \quad .$$

Alors le polynôme

$$r(\boldsymbol{x}) := \frac{g(\boldsymbol{x})}{g(\boldsymbol{s})}$$

est un séparateur de $\boldsymbol{s}$ dans V.

Démonstration. Nous avons $g(\boldsymbol{s}) \neq 0$ et donc $r(\boldsymbol{s}) = 1$. En effet, pour tout $j \in S$ et $l \in E$, $p_j(\boldsymbol{s}) - l = -l \neq 0$ et, l'entier p étant premier, tout produit $\prod_{l \in E} -l$ ne peut s'annuler dans k puisque $E \subset \{1, \ldots, p-1\}$. Soit maintenant $\boldsymbol{t} \in V \backslash \{\boldsymbol{s}\}$, alors il existe $j \in [1, n]$ tel que $p_j(\boldsymbol{t}) \neq 0$; par définition de E, $l = p_j(\boldsymbol{t}) \in E$; puisque le facteur $p_j(\boldsymbol{x}) - l$ de g s'annule en $\boldsymbol{x} = \boldsymbol{t}$, nous parvenons à l'identité $r(\boldsymbol{t}) = g(\boldsymbol{t}) = 0$; ce qui termine la démonstration. □

Proposition 1. *La complexité du calcul proposé dans le théorème 1 est de l'ordre de $o(n.(p-1))$.*

Démonstration. En effet, au pire des cas, $\#S = n$ et $\#E = p-1$. Aussi, $p_j(x) = x_j - s_j$ et le calcul d'un séparateur $r(x)$ va nécessité $2n.(p-1)$ additions, $(n-1).(p-1)$ multiplications et 1 division dans $\mathbb{Z}/p\mathbb{Z}$. □

Remarque 1. Afin d'obtenir un système dynamique polynomial associé à une trajectoire discrète de longueur m, il suffit de calculer $m-1$ séparateurs ; si le résultat n'est pas optimal (voir théorème 2), la complexité du calcul demeure linéaire en n, ce qui est fort intéressant pour trouver un premier modèle d'étude ; les données biologiques étant telles que p et m sont négligeables devant n.

Remarque 2. Le calcul d'un premier SDP peut s'effectuer de manière parallèle, ce qui permet une exécution rapide du programme en dépit de la taille importante de n. En effet, l'exécution simultanée du calcul des séparateurs donnés par le théorème 2 sur $m-1$ processeurs permet un gain en temps et en mémoire.

Algorithme 1. Cet algorithme exécute de manière parallèle le calcul des modules fondamentaux ainsi que celui des séparateurs qui interviennent dans la formule du SDP.

Entrée : n, $m-1$ processeurs et $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{m-1}, \boldsymbol{s}_m$,
Sortie : un SDP $f = (f_1, \ldots, f_n)$ vérifiant $f_j(s_i) = s_{i+1,j}$.

Début

Pour chaque processeur i, calculer :

$$\boldsymbol{p}_i = (p_{i,1} = x_1 - s_{i,1}, \ldots, p_{i,n} = x_n - s_{i,n})$$

$$E_i = \{p_{i,j}(t) \neq 0, j \in S, t \neq \boldsymbol{s}_i\}$$

$$g_i(x) = \prod_{j \in S} \prod_{l \in E_i} (p_{i,j}(x) - l)$$

$$r_i(x) = \frac{g_i(x)}{g_i(\boldsymbol{s}_i)}.$$

Pour j allant de 1 à n faire (à partager sur les $m-1$ processeurs) :

$$f_j(x) = \sum_{i=1}^{m-1} r_i(x) s_{i+1,j}$$

Fin.

Que ce soit en mémoire partagée ou en mémoire distribuée, l'algorithme est efficace. Aussi, il n'y a pas besoin d'échanges entre les processeurs même à la dernière étape de calculs de f et nous avons une préférence pour la technique OpenMP de telle sorte que les processeurs aient accès à la même mémoire (en MPI par exemple, l'algorithme reste aussi efficace).
L'algorithme est donc parfaitement parallélisable et si on note $T(1) = o(n(m-1)(p-1))$ le temps d'exécution sur 1 processeur, alors $T(m-1) = T(1)/m - 1 = o(n(p-1))$, l'accélération $A(m-1) = m-1$ et l'efficacité est de 100%.

C'est dans [12] que R. Thomas a présenté les premiers modèles booléens en biologie. Depuis, plusieurs travaux ont été apportés dont notamment les contributions de [10]. Nous proposons dans la section 2.1 une adaptation des séparateurs algébriques au problème biologique booléen. Dans la section 2.2, nous introduisons un théorème permettant de calculer des séparateurs algébriques de degré optimal (et ceci en enlevant les redondances du théorème 1).

2.1. **Système dynamique booléen.** Supposons, pour simplifier, que l'ensemble séparateur S de V est $\{1, \ldots, n\}$. Dans le cas $p = 2$, E est réduit à $\{1\}$ et les séparateurs sont sous une forme plus compacte. En effet, soient $k = \mathbb{Z}/p\mathbb{Z}$ et les polynômes de $k[x_1, \ldots, x_n]$ donnés par

$$\begin{aligned} q &= \sum_{r=1}^{n} e_r(\boldsymbol{p}) \quad \text{et} \\ r &= q+1 \end{aligned}$$

où $e_1(\boldsymbol{p}), \ldots, e_n(\boldsymbol{p})$ sont les fonctions symétriques élémentaires en les éléments de $\boldsymbol{p} = (p_1, \ldots, p_n)$ (utiliser [13] pour les calculer). Alors pour tout $\boldsymbol{t} \in \mathcal{E}$

$$\begin{aligned} q(\boldsymbol{t}) = 0 \quad &\Leftrightarrow \quad \boldsymbol{t} = \boldsymbol{s} \\ r(\boldsymbol{t}) = 0 \quad &\Leftrightarrow \quad \boldsymbol{t} \neq \boldsymbol{s} \quad . \end{aligned}$$

En particulier, $q(\boldsymbol{t}) = 1 \quad \forall \boldsymbol{t} \neq \boldsymbol{s}$ et $r(\boldsymbol{s}) = 0$. Pour constater ces propriétés sur q et r, considérons le polynôme g univarié en y, à coefficients dans $k[x_1, \ldots, x_n]$ dont les racines sont les modules fondamentaux, s'exprimant ainsi :

$$g(y)(\boldsymbol{x}) = \prod_{j=1}^{n} (y - p_j(\boldsymbol{x})) \quad .$$

Le polynôme $g(1)(x)$ est celui du théorème 1. Pour tout $t \in \mathcal{E}$, nous avons la suite d'équivalences :

$$g(1)(\boldsymbol{t}) = 0 \Leftrightarrow \exists j \in [1, n] \ : p_j(\boldsymbol{t}) = 1 \Leftrightarrow \boldsymbol{t} \neq \boldsymbol{s} \quad .$$

Donc $g(1)$ est un polynôme séparateur pour $\boldsymbol{s}$ dans $\mathcal{E}$. Par ailleurs, d'après l'identité suivante sur les coefficients des polynômes univariés :

$$g(y) = y^n - e_1(\boldsymbol{p}) \cdot y^{n-1} + e_2(\boldsymbol{p}) \cdot y^{n-2} \cdots + (-1)^n \cdot e_n(\boldsymbol{p}) \quad ,$$

nous avons $r = g(1) = q + 1$ dans $\mathbb{Z}/2\mathbb{Z}$.

Exemple 2.1. Soient $m = 3$, $p = 2$, $\boldsymbol{s}_1 = (0, 1, 1)$, $\boldsymbol{s}_2 = (1, 1, 0)$ et $\boldsymbol{s}_3 = (0, 1, 1)$. Les ensembles de module fondamentaux associés à chaque état $\boldsymbol{s}_i$ étant noté $\boldsymbol{p}_i$, nous obtenons :

$$\boldsymbol{p}_1 = \{x_1, x_2 + 1, x_3 + 1\} \quad et \quad \boldsymbol{p}_2 = \{x_1 + 1, x_2 + 1, x_3\}.$$

Les fonctions symétriques élémentaires de $\boldsymbol{p}_1$ sont :
$e_1(\boldsymbol{p}_1) = x_1 + x_2 + x_3$, $e_2(\boldsymbol{p}_1) = x_1x_2 + x_1x_3 + x_2x_3 + x_2 + x_3 + 1$ et $e_3(\boldsymbol{p}_1) = x_1x_2x_3 + x_1x_2 + x_1x_3 + x_1$; donc le séparateur de $\boldsymbol{s}_1$ est

$$r_1(x_1, x_2, x_3) = x_1x_2x_3 + x_2x_3$$

De même, nous trouvons le séparateur r_2 de $\boldsymbol{s}_2$:

$$r_2(x_1, x_2, x_3) = x_1x_2x_3 + x_1x_2 \quad .$$

Pour $t_1 = (1, 0)$, $t_2 = (1, 1)$ et $t_3 = (0, 1)$, nous obtenons $f_1 = r_1, f_2 = r_1 + r_2$ et $f_3 = r_2$. Ainsi, $f = (f_1, f_2, f_3)$ est définie par :

$$f_1 = x_1x_2x_3 + x_2x_3 \quad , \quad f_2 = x_1x_2 + x_2x_3 \quad \text{et} \quad f_3 = x_1x_2x_3 + x_1x_2. \tag{5}$$

Le support de chacune des composantes du système dynamique polynomial f permet de tracer le graphe de dépendance (donné dans la figure 2.1). Notons que sur Cocoa, il nous faut considérer les états $\boldsymbol{s}_1$ et $\boldsymbol{s}_2$ qui donnent $r_1 = x_3$ et $r_2 = x_2 + x_3$. D'où, en tenant compte de $\boldsymbol{s}_3$: $f = (x_3, x_2, x_2 + x_3)$. Ce système dynamique polynomial (booléen) nous permet d'obtenir le graphe de dépendance[3] de la figure 3 (graphe qui comprend un nombre plus réduit d'arcs qu'avec le système dynamique polynomial (5)).

3. Les figures ont été obtenues par le logiciel DVD de [6]

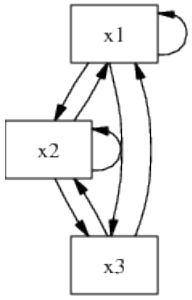

FIGURE 2. *Graphe des intéraction de* $f = (x_1x_2x_3 + x_2x_3, x_1x_2 + x_2x_3, x_1x_2x_3 + x_1x_2)$ *pour* $n = 3$, $m = 3$ *et* $p = 2$

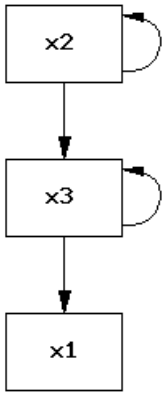

FIGURE 3. *Graphe des intéractions de* $f = (x_3, x_2, x_2 + x_3)$ *pour* $n = 3$, $m = 3$ *et* $p = 2$

2.2. **Séparateur algébrique optimal.** Dans le théorème 1, nous séparons plusieurs fois le point $\boldsymbol{s}$ d'un même point $\boldsymbol{t}$ de V. Nous réduisons donc le polynôme g de la façon suivante :

Définitions 2.2.

– L'*ensemble séparateur de* $\boldsymbol{s}$ *et* $\boldsymbol{t}$ deux points distincts de V :

$$S(\boldsymbol{s}, \boldsymbol{t}) = \{j \in S \mid s_j \neq t_j\} = \{j \in S \mid p_j(\boldsymbol{t}) \neq 0\} \quad ,$$

où S est l'ensemble séparateur de V ;

– L'*ensemble initial des polynômes séparateurs de* $\boldsymbol{s}$ *et* $\boldsymbol{t}$

$$PS(\boldsymbol{s}, \boldsymbol{t}) = \{p_{\boldsymbol{t},j}(\boldsymbol{x}) \mid j \in S(\boldsymbol{s}, \boldsymbol{t})\} \quad ;$$

où $p_{\boldsymbol{t},j}(\boldsymbol{x}) = x_j - t_j$;
pour tout $g \in PS(\boldsymbol{s}, \boldsymbol{t})$, puisque $j \in S(\boldsymbol{s}, \boldsymbol{t})$, nous avons $g(\boldsymbol{s}) = -p_{\boldsymbol{t},j}(\boldsymbol{s})$ ce qui implique que $g(\boldsymbol{t}) = 0$; tout produit de séparateur initial de $\boldsymbol{s}$ et $\boldsymbol{t}$ est un séparateur de ces deux points.

– Un *ensemble initial minimal séparant* $\boldsymbol{s}$ *dans* V, noté $Min(\boldsymbol{s}, V)$ est composé d'au plus $m - 1$ polynômes (distincts) g tels que pour chaque $\boldsymbol{t} \in V$ distinct de $\boldsymbol{s}$, il existe un et un seul élément

de $Min(\boldsymbol{s}, V)$ appartenant à $PS(\boldsymbol{s}, \boldsymbol{t})$; i.e. pour chaque $\boldsymbol{t}$ dans V distinct de $\boldsymbol{s}$, l'intersection de $Min(\boldsymbol{s}, V)$ avec $PS(\boldsymbol{s}, \boldsymbol{t})$ est réduite à un et un seul élément.

Exemple 2.3. Pour illustrer les définitions précédentes, considérons les états $\boldsymbol{s}_1 = (2, 1, 2)$, $\boldsymbol{s}_2 = (1, 1, 0)$, $\boldsymbol{s}_3 = (0, 0, 1)$ et $\boldsymbol{s}_4 = (1, 2, 0)$. Nous avons $S = \{1, 2, 3\}$ et

$$S(\boldsymbol{s}_1, \boldsymbol{s}_2) = \{1, 3\} \quad S(\boldsymbol{s}_2, \boldsymbol{s}_4) = \{2\} \quad S(\boldsymbol{s}_1, \boldsymbol{s}_3) = S(\boldsymbol{s}_1, \boldsymbol{s}_4) = S(\boldsymbol{s}_2, \boldsymbol{s}_3) = S(\boldsymbol{s}_3, \boldsymbol{s}_4) = \{1, 2, 3\}$$

En mettant les polynômes sous une forme normalisée permettant d'identifier ceux qui donnent le même résultat (même dénominateur, division par le pgcd des coefficients, signe du coefficient de la plus petite variable), nous obtenons pour $\boldsymbol{s}_1$:

$$PS(\boldsymbol{s}_1, \boldsymbol{s}_2) = \{x_1 - 1, x_3\} \quad PS(\boldsymbol{s}_1, \boldsymbol{s}_3) = \{x_1, x_2, x_3 - 1\} \quad PS(\boldsymbol{s}_1, \boldsymbol{s}_4) = \{x_1 - 1, x_2 - 2, x_3\}.$$

Voici quelques ensembles initiaux minimaux de $\boldsymbol{s}_1$ dans V (on peut tous les énumérer dans un cadre fini) :

$$\{x_1 - 1, x_1\} \quad \{x_1 - 1, x_2\} \quad \{x_3, x_2\}.$$

Les ensembles initiaux des polynômes séparateurs de $\boldsymbol{s}_4$ sont

$$PS(\boldsymbol{s}_4, \boldsymbol{s}_1) = \{x_1 - 2, x_2 - 1, x_3 - 2\} \quad PS(\boldsymbol{s}_4, \boldsymbol{s}_2) = \{x_2 - 1\} \quad PS(\boldsymbol{s}_4, \boldsymbol{s}_3) = \{x_1, x_2, x_3 - 1\}.$$

Théorème 2. *Soit $Min(\boldsymbol{s}, V)$ un ensemble minimal séparant $\boldsymbol{s}$ dans V, et soit*

$$\mathcal{G} = \prod_{g \in Min(\boldsymbol{s}, V)} g$$

alors

$$r(\boldsymbol{x}) = \frac{\mathcal{G}(\boldsymbol{x})}{\mathcal{G}(\boldsymbol{s})}$$

est un séparateur pour $\boldsymbol{s}$ dans V.

Exemple 2.4. Poursuivons l'exemple 2.3 et calculons un séparateur r_1 de $\boldsymbol{s}_1 = (2, 1, 2)$. A partir de l'ensemble minimal $\{x_3, x_2\}$ de $\boldsymbol{s}_1$ dans V, nous obtenons le polynôme

$$r_1 = \frac{x_2 x_3}{2} = -x_2 x_3$$

valant 1 en $\boldsymbol{s}_1$ et 0 ailleurs dans V. C'est le même que celui trouvé par Cocoa (voir section 4). Un séparateur de s_4 obtenu par le théorème 2 est par exemple

$$r_4 = (x_1 - 2)(x_2 - 1)(x_3 - 1) = x_1 x_2 x_3 - x_1 x_2 - x_1 x_3 + x_2 x_3 + x_1 - x_2 - x_3 + 1.$$

Le calcul des SDP est parallélisable et la complexité parallèle est intéressante sauf pour calculer les $S(\boldsymbol{s}, \boldsymbol{t})$. En effet, si nous voulons paralleliser cette partie, il faut calculer $S(\boldsymbol{s}, \boldsymbol{t})$ et $S(\boldsymbol{t}, \boldsymbol{s})$: on n'en calcule que $m - 1$ à chaque fois alors qu'en non-parallèle (en séquentiel) on en calcule $(m - 1)!$. La parallélisation est détaillée dans l'algorithme suivant :

Algorithme 2. Cet algorithme exécute de manière parallèle le calcul des ensembles séparateurs ainsi que celui des séparateurs qui interviennent dans la formule des SDP obtenus par le Théorème 2.

Branche pour $\boldsymbol{s}_1$:	temps 1 :	Attente de réception
	temps 2 :	Réception de $S(\boldsymbol{s}_2, \boldsymbol{s}_1), \dots S(\boldsymbol{s}_m, \boldsymbol{s}_1)$ des autres cellules et Calcul du séparateur r_1 de $\boldsymbol{s}_1$
Branche pour $\boldsymbol{s}_2$:	temps 1 :	Calcul de $S(\boldsymbol{s}_2, \boldsymbol{s}_1)$ et envoi à Branche pour $\boldsymbol{s}_1$
	temps 2 :	Attente de réception
	temps 3 :	Réception de $S(\boldsymbol{s}_3, \boldsymbol{s}_2)$, ..., $S(\boldsymbol{s}_{m-1}, \boldsymbol{s}_2)$ et Calcul du sépareteur r_2 de $\boldsymbol{s}_2$
		...
Branche pour $\boldsymbol{s}_i$:	temps j < i :	Calcule de $S(\boldsymbol{s}_i, \boldsymbol{s}_j)$ et envoi à Branche $\boldsymbol{s}_j$
	temps j :	Attente de réception
	temps j+1 :	Réception de $S(\boldsymbol{s}_{i+1}, \boldsymbol{s}_i)$,..., $S(\boldsymbol{s}_{m-1}, \boldsymbol{s}_i)$ et Calcul du sépareteur r_i de $\boldsymbol{s}_i$
		...
Branche pour $\boldsymbol{s}_{m-1}$:	temps $j = 1$ à $m-1$	Calcul de $S(\boldsymbol{s}_{m-1}, \boldsymbol{s}_j)$ et Envoi à Branche $\boldsymbol{s}_j$ et Calcul de r_m

Notons enfin que les séparateurs obtenus par le théorème 2 sont de degré $\leq n$ alors que ceux obtenus par le théorème 1 sont de degré $\leq n.\#E$. Par exemple, pour $n = 3$ et $E = \{1, 2\}$, $n.\#E = 6$ et le séparateur de s_1 obtenu par le théorème 1 (donc avec une complexité linéaire) est

$$r_1 = -x_1(x_1 - 1)(x_2 - 2)x_2x_3(x_3 - 1).$$

3. Séparateurs Affines

Nous détaillons dans cette section une méthode permettant de déterminer, dans certains cas, des séparateurs affines. Notons

$$CL(\boldsymbol{t})$$

l'ensemble des combinaisons linéaires des générateurs (linéaires ici) de $\mathfrak{M}_{\boldsymbol{t}}$, l'idéal des $\boldsymbol{t}$-relations ; afin de ne pas avoir plusieurs fois la "même" relation, nous choisissons une forme normalisée pour chaque polynôme (par exemple, on met le polynôme au même dénominateur (que l'on retire), on le divise par son contenu (le pgcd de ses coefficients) et on multiplie le coefficient de la plus petite variable par son signe afin de le rendre positif (par exemple, $x_1 < x_2 < \cdots < x_n$; sachant que les auteurs de [8] expliquent que l'ordre des x_i peut être préalablement déterminé par des informations concernant l'influence de certains produits de gènes sur d'autres).

Pour trouver les séparateurs affines de $\boldsymbol{s}$ dans V, il faut déterminer

$$(\cap_{\boldsymbol{t} \neq \boldsymbol{s}} CL(\boldsymbol{t})) \backslash CL(\boldsymbol{s}) \qquad .$$

cet ensemble est un ensemble affine de séparateurs de $\boldsymbol{s}$ dans V (avec le produit c'est peut-être algébrique mais efficace à calculer si on les veut tous ...). Pour savoir si une relation n'appartient pas à $\mathfrak{M}_{\boldsymbol{s}}$ (i.e. à $CL(\boldsymbol{s})$), il suffit de la diviser en les modules fondamentaux de $\mathfrak{M}_{\boldsymbol{s}}$ (ou ce qui revient au même à l'évaluer en $\boldsymbol{s}$). D'une façon générale

$$r_j \in \cap_{i\neq j}\mathfrak{M}_{\boldsymbol{s}_i} \quad \text{et} \quad r_j \notin \mathfrak{M}_{\boldsymbol{s}_j}$$

pour tout séparateur r_j de $\boldsymbol{s}_j$ dans V.

Exemple 3.1. $\boldsymbol{s}_1 = (2,1,2)$, $\boldsymbol{s}_2 = (4,4,0)$, $\boldsymbol{s}_3 = (3,3,1)$, $\boldsymbol{s}_4 = (1,2,3)$; $S = \{1,2,3,4\}$. Nous additionnons, par exemple, d'abord les générateurs 2-à-2 $(p_{i,j}+p_{i,k})$ puis les soustrait $(p_{i,j}-p_{i,k})$, puis $2p_{i,j}+p_{i,k}$, etc ... puis on passe à $p_{i,1}+p_{i,2}+p_{i,3}$; Par exemple

$$\begin{aligned}
\mathfrak{M}_{\boldsymbol{s}_1} &= < x_1+3, x_2+4, x_3+3 > \\
CL(\boldsymbol{s}_1) &= \{0, 2x_1+3x_2+x_3+1, 3x_1+2x_2+3x_3+1, 3x_1-x_2+x_3+3, \ldots, \\
&\quad x_2-x_3+1, x_1+x_2+2x_3+3, 3x_1+2x_2+2, \ldots, \mathbf{2x_1+2x_2-1}, \ldots\} \\
\mathfrak{M}_{\boldsymbol{s}_2} &= < x_1+1, x_2+1, x_3 > \\
CL(\boldsymbol{s}_2) &= \{-x_2+4, -x_1-x_2-x_3+3, 3x_1+2x_2+3x_3, 3x_1-x_2+x_3+2, \\
&\quad -x_1-x_3+4, 3x_1+x_3+3, 2x_1+x_2+x_3+3, x_1+x_2+3x_3+2, \\
&\quad 3x_1+x_2+2x_3+4, x_1+3x_2+x_3+4 \quad \mathbf{2x_1+2x_2-1}, 2x_1+2x_2+3x_3+4, \ldots\} \\
\mathfrak{M}_{\boldsymbol{s}_3} &= < x_1+2, x_2+2, x_3+4 > \\
CL(\boldsymbol{s}_3) &= \{-x_2+3, 3x_1+3x_2+2, 3x_1+2x_2+3x_3+2, 3x_1-x_2+x_3+3, \\
&\quad \ldots, x_1-x_2+3x_3+2, x_1+3x_3+4\} \\
\mathfrak{M}_{\boldsymbol{s}_4} &= < x_1+4, x_2+3, x_3+2 > \\
CL(\boldsymbol{s}_4) &= \{-x_1-x_2-x_3+1, 3x_1-x_2+x_3+1, -x_1-x_3+4, 3x_1+2x_2+3x_3+4, \\
&\quad \mathbf{2x_1+2x_2-1}, x_1+3x_3, 3x_1+3x_2+x_3+3, -x_2+2\ldots\}
\end{aligned}$$

Nous trouvons rapidement le séparateur $2x_1+2x_2-1$ de s_3 que l'on peut aussi retrouver par une identification linéaire. $r_3 = 2x_1+2x_2-1$ est un élément de $(\cap_{i\neq 3}CL(\boldsymbol{s}_i))$ qui n'appartient pas à $CL(\boldsymbol{s}_3)$. Dans cet exemple il n'est pas possible de trouver des séparateurs affines de s_1, s_2 et s_4.

Remarque 3. Ce qui est intéressant dans cette méthode, c'est qu'elle peut être traitée entièrement dynamiquement. Il est hors de question de calculer $CL(\boldsymbol{t})$ en entier ; là, il y a un algorithme intéressant à mettre en place : plusieurs stratégies sont possibles notamment en calcul parallèle.

4. Exemple pour $n = 3$, $p = 3$ et $m = 5$

En partant de 3 produits de gènes x_1, x_2 et x_3 appartenant à $\mathbb{Z}/3\mathbb{Z}$ et la séquence de 5 temps suivantes : $\boldsymbol{s}_1 = (2,1,2); \boldsymbol{s}_2 = (1,1,0); \boldsymbol{s}_3 = (0,0,1); \boldsymbol{s}_4 = (1,2,0)$ et $\boldsymbol{s}_5 = (0,0,1)$, nous obtenons avec le méthode de Laubenbacher et Stigler que les séparateurs de l'espace projectif en les points $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ et $\boldsymbol{s}_4$ sont respectivement :

$$r_1(x_1,x_2,x_3) = -\mathbf{x_2x_3},$$
$$r_2(x_1,x_2,x_3) = -\mathbf{x_1} - \mathbf{x_2},$$
$$r_3(x_1,x_2,x_3) = x_2x_3 + x_3^2,$$
$$r_4(x_1,x_2,x_3) = x_1x_2 - x_2^2 + x_2x_3.$$

Partant des modules fondamentaux associés repectivement à $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, $\boldsymbol{s}_3$ et $\boldsymbol{s}_4$:

$$\boldsymbol{p}_1 = \{x_1 + 1, x_2 - 1, x_3 + 1\},$$
$$\boldsymbol{p}_2 = \{x_1 - 1, x_2 - 1, x_3\},$$
$$\boldsymbol{p}_3 = \{x_1, x_2, x_3 - 1\},$$
$$\boldsymbol{p}_4 = \{x_1 - 1, x_2 + 1, x_3\}.$$

Nous obtenons avec la méthode algébrique la même chose pour les deux premiers séparateurs :

$$r_1(x_1, x_2, x_3) = -\mathbf{x_2 x_3},$$
$$r_2(x_1, x_2, x_3) = -\mathbf{x_1} - \mathbf{x_2},$$
$$r_3(x_1, x_2, x_3) = x_1 x_3 + x_3,$$
$$r_4(x_1, x_2, x_3) = -x_2^2 + x_2.$$

Notons que nous obtenons beaucoup d'autres séparateurs grâce à la méthode développée à la section 2.2. Par exemple, des séparateurs de s_3 dans V sont $\{-x_1^2 + 1, -x_2x_3 + x_3, x_1x_3 + x_1 - x_3 + 2, x_1x_3 + x_3, -x_2^2 + 1, x_1x_2 - x_1 - x_2 + 1, -x_3^2 - x_3, \ldots\}$ et des séparateurs de s_4 dans V sont $\{x_1x_2 - x_1, -x_2x_3 + x_2 + x_3 + 2, \ldots\}$.

D'après la méthode de la section 3, nous pouvons également affirmer qu'il n'existe pas de séparateur affine de $s4$. Aussi, nous obtenons plusieurs candidats possibles aux Systèmes dynamiques polynomiaux dont notamment

$$f = (x_1x_3 - x_2x_3 + x_3, -x_1x_3 - x_2x_3 - x_3, -x_2^2 - x_1).$$

Le graphe des intéractions est donné dans la figure 4 et le graphe des états de transitions (figure 5) présente un point fixe (0,0,0) et deux cycles (de longueur 2 et 5) sont obtenus.

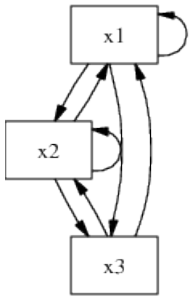

FIGURE 4. *Graphe des états de transition ($n = 3$, $m = 3$ et $p = 5$) de $f = (x_1x_3 - x_2x_3 + x_3, -x_1x_3 - x_2x_3 - x_3, -x_2^2 - x_1)$.*

Le système dynamique polynomial obtenu par la méthode de Stigler et Laubenbacher est égal à $(x_3^2, x_2x_3 - x_3^2, -x_2^2 + x_2x_3 - x_2)$. Cela donne la figure 6 pour le graphe d'intéractions (le graphe de transition ne présente qu'un point fixe (0,0,0), voir [9]).

Notons enfin que, plus que le degré, c'est le nombre de variables apparaissant dans chaque composante du SDP qui fera la différence et ce qui compte, c'est d'avoir un modèle qui colle le plus à la

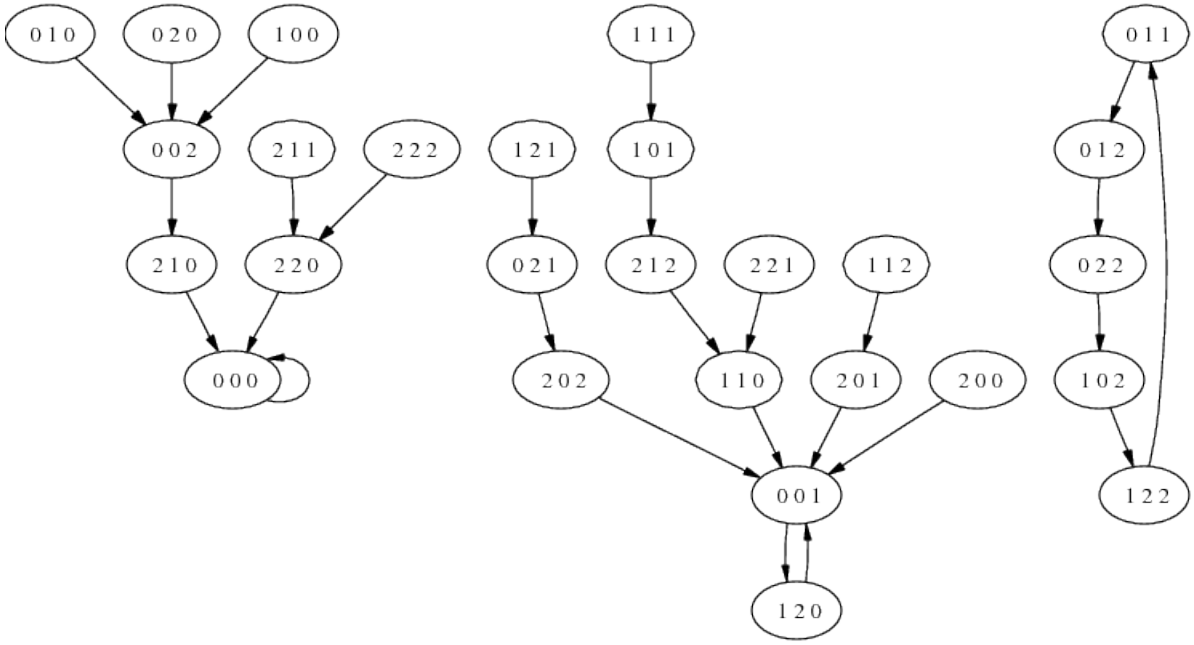

FIGURE 5. *Graphe des intéractions ($n = 3$, $m = 3$ et $p = 5$) de* $f = (x_1x_3 - x_2x_3 + x_3, -x_1x_3 - x_2x_3 - x_3, -x_2^2 - x_1)$.

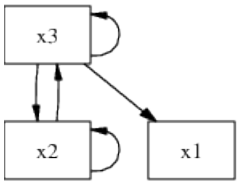

FIGURE 6. *Graphe des intéractions ($n = 3$, $m = 3$ et $p = 5$) de* $f = (x_3^2, x_2x_3 - x_3^2, -x_2^2 + x_2x_3 - x_2)$

réalité des donnés (voir [6]).

5. Conclusion et Perspectives

Notre approche s'inscrit dans un cadre plus global d'une équipe pluridisciplinaire travaillant sur une modélisation aussi bien algébrique que se basant sur des EDO ou des réseaux bayésiens. C'est dans cette optique que nous avons commencé dans [1] par présenter une synthèse des principales méthodes s'adaptant à ce cadre. La méthode algébrique présenté dans cet article adapte des techniques d'interpolation de Lagrange et de théorie de Galois à des problématiques de bio-systémique. Cette approche permet de modéliser efficacement des données biologiques de taille imortante avec

des outils relativement simples : le calcul de polynômes symétriques élémentaires dans le cas booléens ou celui d'ensembles initiaux minimaux dans le cas polynomial.

Ainsi, nous obtenons un premier modèle sous forme d'un SDP avec une complexité linéaire de l'ordre de $o(n.(p-1).(m-1))$. Beaucoup d'autres SDP peuvent être déterminés avec notre approche de manière optimale. Reste qu'obtenir tous les systèmes dynamiques polynomiaux à partir d'un système dynamique polynomial calculé algébriquement est également possible : un calcul des générateurs de l'idéal $I(V) = \{h \in k[\boldsymbol{x}] \quad | \, h(\boldsymbol{s}) = 0 \quad , \forall \boldsymbol{s} \in V\}$ permet par exemple d'effectuer ce calcul (voir [9]). Le calcul de tous les SDP passe par la détermination de tous les séparateurs : un séparateur $r(\boldsymbol{x})$ de $\boldsymbol{s}$ dans V résulte du choix d'un élément de l'ensemble

$$\{r \in \cap_{\boldsymbol{t} \neq \boldsymbol{s}} \mathfrak{M}_{\boldsymbol{t}} \quad \text{et} \quad r \notin \mathfrak{M}_{\boldsymbol{s}} \quad , \forall \boldsymbol{t} \in V\}.$$

Dans [2], les auteurs développent un algorithme combinatoire permettant de générer une base de monômes linéaires. Cet algorithme, facile à implémenter et parallélisable, sera utilisé dans le calcul de générateurs de l'idéal $I(V)$.

Références

[1] Ines Abdeljaoued-Tej, Alia Ben Kahla, Slimane Ben Miled, Lamia Guizani-Tabbane, and Koussay Dellagi. *Handbook of Research on Systems Biology Application in Medicine*, chapter System biology and infectious diseases. Editor : Andriani Daskalaki, Max Planck Institute, Berlin, 2009.

[2] L. Cerlienco and M. Mureddu. From algebraic sets to monomial linear bases by means of combinatorial algorithms. In *Proceedings of the 4th conference on Formal power series and algebraic combinatorics*, pages 73–87, Amsterdam, The Netherlands, The Netherlands, 1995. Elsevier Science Publishers B. V.

[3] Elena S. Dimitrova, Abdul Salam Jarrah, Reinhard C. Laubenbacher, and Brandilyn Stigler. A gröbner fan method for biochemical network modeling. In Dongming Wang, editor, *ISSAC*, pages 122–126, 2007.

[4] Elena S. Dimitrova, John J. McGee, and Reinhard C. Laubenbacher. Discretization of time series data, 2005.

[5] Ines Abdeljaoued-Tej, Alia Ben Kahla, Slimane Ben Miled and Lamia Guizani-Tabbane. *Calcul Formel et Applications à la Biologie*. Premier Colloque franco-maghrébin de Calcul Formel, 23-26 mai 2008, Iles de Kerkennah, Tunisie.

[6] Abdul Salam Jarrah, Reinhard Laubenbacher, Brandilyn Stigler, and Michael Stillman. Reverse-engineering of polynomial dynamical systems, 2006.

[7] Joseph-Louis Lagrange. Réflexions sur la résolution algébrique des équations. *Prussian Academy*, 1770.

[8] Reinhard Laubenbacher. *Modeling and simulation in Applied Mathematics*, volume 64 of *Modeling and Simulation of Biological Networks : American Mathematical Society, Short Course*. AMS Bookstore, 2007.

[9] Reinhard Laubenbacher and Brandilyn Stigler. A computational algebra approach to the reverse engineering of gene regulatory networks. *Journal of Theoretical Biology*, 229 :523, 2004.

[10] Reinhard C. Laubenbacher. A computer algebra approach to biological systems. In *ISSAC*, pages 5–6, 2003.

[11] N. Tchebotarev. *Gründzüge des Galois'shen Theorie*. P. Noordhoff, 1950.

[12] Rene Thomas. *Kinetic logic : a boolean approach to the analysis of complex regulatory systems*, volume 29. Lecture Notes in Biomathematics, 1979.

[13] Annick Valibouze. Symbolic computation with symmetric polynomials, an extension to Macsyma. In *Computers and Mathematics (MIT, USA, June 13-17, 1989)*, pages 308–320. Springer-Verlag, New York Berlin, 1989. (Voir "Symmetries" dans `http://maxima.sourceforge.net/docs.shtml`).

[14] Annick Valibouze. Dépendance algébrique des zéros de polynômes et groupes de Galois. *Bull. Math. Soc. Sci. Math. Roumanie (N.S.)*, 48(96)(1) :73–96, 2005.