

(public 2010)

Résumé : Modélisation d'un tissu cellulaire lors d'une contamination.

Mots clefs : Automates, Algorithme et Réécriture

- *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. Vous êtes laissé(e) libre d'organiser votre discussion comme vous l'entendez. Des suggestions de développement, largement indépendantes les unes des autres, vous sont proposées en fin de texte. Vous n'êtes pas tenu(e) de les suivre. Il vous est conseillé de mettre en lumière vos connaissances à partir du fil conducteur constitué par le texte. Le jury appréciera que la discussion soit accompagnée d'exemples traités sur ordinateur.*

1. Introduction

Les organismes vivants sont composés d'organes et de tissus eux-mêmes composés de cellules.

Le fonctionnement des différents organes et tissus est très étudié depuis plusieurs siècles (biologie et médecine). Celui de la cellule seule l'est depuis plusieurs décennies (biologie cellulaire). Par contre, le niveau intermédiaire, l'agrégat de cellules, très complexe, n'est pas étudié aussi finement. Le besoin d'expérimentation *in silico* est très important et les enjeux énormes : « fabrication » de tissus de remplacement (par exemple peau pour les grands brûlés), compréhension des mécanismes de régénération (rein, foie), amélioration de la croissance des végétaux... La simulation aide à imaginer de meilleurs traitements, plus rapides et plus sûrs.

Il existe une très grande variété parmi les cellules : globules rouges, neurones, cellules du muscle, du foie, de l'estomac... Cette variété correspond aux différents rôles des cellules et se retrouve selon différents critères : la taille et l'imbrication (par exemple neurones par rapport aux cellules musculaires), la mobilité (les globules rouges se déplacent dans le système sanguin), le mode de fonctionnement (traiter des influx électriques pour les cellules nerveuses, accumulation, transformation, stockage et libération de substances chimiques dans différentes glandes, contraction pour les cellules musculaires).

Actuellement, on imagine difficilement modéliser et simuler sérieusement un organe entier ou un organisme vivant même composé de relativement peu de cellules. Par contre, on peut imaginer modéliser un tissu cellulaire de taille raisonnable tant que les cellules sont relativement homogènes et que leur comportement élémentaire n'est pas trop complexe.

On se propose donc de modéliser la propagation d'une infection dans un tissu. On suppose le tissu allongé et peu épais, au point de considérer toutes les cellules placées les unes à côté des autres sur une ligne. Au départ une ou plusieurs cellules sont infectées. Une cellule infectée

passer par différents stades au cours de son évolution. À certains stades de l'infection, elle peut infecter les cellules voisines.

Dans cette épreuve, il n'est bien évidemment pas demandé une connaissance approfondie des mécanismes biologiques, mais juste de faire des hypothèses raisonnables et suffisamment simplificatrices pour présenter une modélisation dans le temps imparti. Le programme proposé n'a aucune prétention de réalisme.

2. Aspect algorithmique

Exercice de programmation :

- *Il vous est demandé de rédiger un programme conforme aux spécifications ci-dessous dans l'un des langages C, Caml ou Java à votre choix. Ce programme devra être accompagné d'un exemple d'exécution permettant d'en vérifier le bon fonctionnement. La clarté et la concision du programme seront des éléments importants d'appréciation pour le jury.*

On veut simuler un phénomène de contamination. L'état d'une cellule est représenté par un entier compris entre 0 et K (K est un paramètre fixé). Une cellule est normalement saine, non contaminée, dans l'état 0 et reste dans cet état tant qu'elle n'est pas infectée. Tous les autres états représentent différents stades d'infection. Une cellule infectée commence dans l'état 1 et à chaque itération passe dans l'état suivant (c.-à-d. 2 puis 3 puis...). Quand elle atteint l'état K , elle a remporté sa lutte contre l'infection et retourne à l'état 0, saine et non infectée à l'itération suivante.

Une cellule infectée n'est pas tout de suite contagieuse (c.-à-d. capable d'infecter ses voisines). Une cellule devient infectieuse à partir du stade I ($I \leq K$, I est un autre paramètre fixé) et le reste jusqu'à sa guérison (retour à l'état 0). Durant cette période, elle peut contaminer chacune de ses deux voisines (les plus proches à droite et à gauche). Si celles-ci ne sont pas déjà infectées, elles le deviennent automatiquement, au stade 1 ; si elles sont déjà infectées, elles le restent.

Les cellules ne bougent pas et on les suppose rangées les unes à côté des autres sur une ligne. Elles sont toutes mises à jour en même temps, de manière synchrone. Pour simplifier, on suppose que les première et dernière cellules restent toujours dans le même état 0.

Tout ceci est schématisé sur la Fig. 1 où l'on voit en haut une configuration et en bas la configuration suivante. Chaque cellule est mise à jour en fonction de ses deux plus proches voisines, sauf pour les cellules des deux extrémités dont l'état ne change pas. Après la configuration $(0, 0, 3, 0, 0, 1, 5, \dots, 0)$, le système passe donc à la configuration $(0, 1, 4, 1, 0, 2, 0, \dots, 0)$ pour $K = 5$ et $I = 2$. Dans cette illustration, l'état de la troisième cellule (3) est en gras pour indiquer pour quelles cellules il intervient durant la mise à jour.

Écrire un programme permettant d'afficher la progression de la contamination sur une vingtaine d'itérations, par exemple pour les valeurs ($K = 5$ et $I = 2$) en partant d'une seule cellule infectée sur une ligne d'une vingtaine de cellules saines. Préciser la complexité en temps et en espace de la simulation.

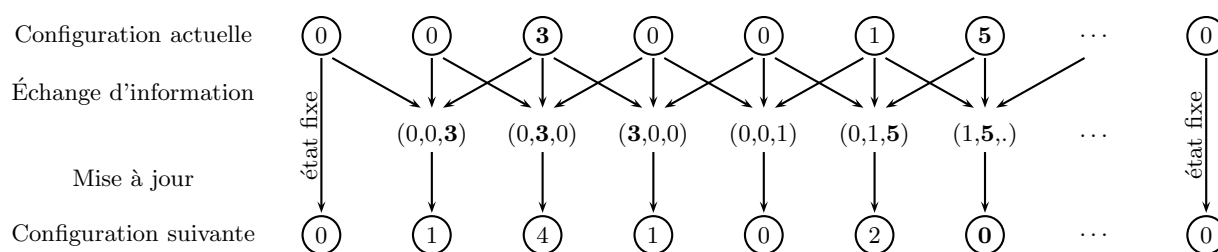


FIG. 1. Modélisation des influences entre cellules ($K = 5$ et $I = 2$).

3. Fonctions continues

On peut considérer du point de vue mathématique les fonctions permettant de passer d'une configuration à la suivante et définissant ainsi la dynamique du système. Une configuration est considérée comme une ligne d'état, soit un élément de $Q^{\mathbb{Z}}$ où Q est un ensemble fini d'états (on suppose qu'il y a au moins 2 états). On considère qu'il y a une infinité de cellules (c.-à-d. \mathbb{Z} et non un intervalle des entiers) pour ne plus avoir de problèmes aux bornes. La fonction de mise à jour $f : Q^3 \rightarrow Q$ permet de définir une fonction globale G de la manière suivante :

$$\forall c \in Q^{\mathbb{Z}}, \forall i \in \mathbb{Z}, (G(c))_i = f(c_{i-1}, c_i, c_{i+1}) .$$

Le nouvel état d'une cellule est toujours calculé à partir des états d'une cellule et de ses deux voisines. Cela correspond à l'illustration de la Fig. 1.

Les fonctions ainsi calculées sont continues pour la topologie produit. Celle-ci se définit de la manière suivante : Q est un ensemble fini, il est naturellement muni de la topologie discrète (c.-à-d. tout sous-ensemble de Q est un ouvert – et un fermé) ; la topologie pour le produit infini $Q^{\mathbb{Z}}$ est alors engendrée par l'ensemble d'ouverts suivant : $Q^{\mathbb{Z}}$ dans lequel on remplace un facteur Q par n'importe quel ouvert de Q (c.-à-d. $Q^{]-\infty, a-1]} \times R \times Q^{[a+1, +\infty[}$ avec a dans \mathbb{Z} et R une partie quelconque de Q). Par exemple l'ensemble des configurations telles que l'état à la position 0 soit un état donné de Q est un ouvert.

Cette topologie vérifie les propriétés suivantes :

- (1) tous les ouverts de base sont aussi des fermés ;
- (2) tous les ouverts sont infinis et indénombrables ;
- (3) il existe pourtant des ouverts qui ne sont pas fermés (et des fermés qui ne sont pas ouverts, p.e. les singletons, ce n'est donc pas la topologie discrète) ; et
- (4) elle est séparée (utiliser une composante où les deux configurations diffèrent).

4. Contamination Probabiliste

Jusqu'ici, l'infection se produit forcément. On peut être un peu plus optimiste et estimer qu'il y a une probabilité constante pour une cellule infectée de contaminer une cellule Saine. On peut

alors chercher à connaître le nombre moyen d'itérations pour que la n^e cellule soit infectée, le nombre moyen de cellules infectées à l'itération t ...

On se place dans le cas suivant : à $t = 0$ seule la cellule numéro 0 est infectée ; dès qu'une cellule est infectée elle reste infectée et contaminante à jamais. Une cellule ne contamine sa voisine qu'avec une probabilité p (on note $q = 1 - p$ la probabilité que l'infection ne soit pas transmise). Le début de l'évolution est dans la table 1 (où l'on note \cdot les cellules saines et X les contaminées).

TAB. 1. Évolution probabiliste.

Itération	Configuration (probabilité)
0	$X \dots (1)$
1	$X \dots (q) \quad XX \dots (p)$
2	$X \dots (q^2) \quad XX \dots (2pq) \quad XXX \dots (p^2)$

On note N_t la variable aléatoire correspondant au nombre de cellules infectées à l'itération t . On a donc au départ : $\mathcal{P}(N_0 = 1) = 1$; après la première itération : $\mathcal{P}(N_1 = 1) = q$, et $\mathcal{P}(N_1 = 2) = p$; et après la deuxième itération : $\mathcal{P}(N_2 = 1) = q^2$, $\mathcal{P}(N_2 = 2) = 2.p.q$ et $\mathcal{P}(N_2 = 3) = p^2$.

On peut aussi se poser la question de savoir si une cellule est ou non infectée. On définit pour cela la variable aléatoire booléenne $I_t(k)$. Celle-ci est vraie si et seulement si la cellule k est infectée à l'itération t . Par exemple, $\mathcal{P}(I_t(0)) = 1$ pour tout t , $\mathcal{P}(I_0(k)) = 0$ pour $0 < k$, $\mathcal{P}(I_1(1)) = p$, $\mathcal{P}(I_2(1)) = p + p.q$, $\mathcal{P}(I_3(1)) = p + p.q + p.q^2$ et $\mathcal{P}(I_2(2)) = p^2$.

On peut prouver la relation de récurrence suivante :

$$(1) \quad \mathcal{P}(I_{t+1}(k+1)) = \mathcal{P}(I_t(k+1)) + p.\mathcal{P}(\neg I_t(k+1) \wedge I_t(k)) .$$

5. Étendre la modélisation

Dans cette partie, on évoque beaucoup de choses que ne peut gérer notre modèle et qui sont des pistes de réflexion pour l'élargir.

Mettre les cellules en ligne n'est pas très réaliste. On peut imaginer d'autres dispositions (maillages réguliers en deux ou trois dimensions, positions quelconques). On peut également imaginer avoir une influence des cellules à distance 2, ou à distance au plus d (cette distance pouvant être celle selon le maillage ou euclidienne si l'on considère des placements quelconques).

Certaines cellules pourraient se déplacer (volontairement ou non).

De nouvelles cellules pourraient apparaître par division, et inversement d'autres pourraient mourir et disparaître. Cela voudrait dire que la position et les connexions des cellules seraient dynamiques.

Suggestions pour le développement

- ▶ *Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.*
 - Démontrer une ou plusieurs des propriétés de la topologie produit sur $\mathcal{Q}^{\mathbb{Z}}$.
 - Montrer, ou argumenter en faveur de, la continuité des fonctions globales ainsi définies.
 - Donner une expression simple de $\mathcal{P}(N_t = k)$, de $\mathcal{P}(I_t(1))$ et de $\mathcal{P}(I_t(t))$ (définie dans la Sect. 5).
 - Établir et simplifier (1) (en interprétant $\neg I_t(n+1) \wedge I_t(n)$).
 - Donner une expression simple de $\mathcal{P}(I_t(2))$ et de $\mathcal{P}(I_{k+1}(k))$ ainsi qu'une formule non récursive pour $\mathcal{P}(I_t(k))$.
 - Si l'on veut étendre la modélisation pour couvrir une ou plusieurs pistes de la Sect. 5, quelles sont les structures de données que l'on pourrait utiliser ?
 - Si l'on passe à des voisinages variables, quelles sont les implications sur la modélisation des dépendances entre cellules d'une part et sur les fonctions de transition d'autre part ?
 - Si très peu de cellules sont infectées, il n'est pas nécessaire de mettre à jour toutes les cellules. Quelles solutions algorithmiques pourrait-on proposer pour limiter les mises à jour inutiles ?