



RECHERCHE OPÉRATIONNELLE ET DÉVELOPPEMENT DURABLE

MASTER PARISIEN DE RECHERCHE OPÉRATIONNELLE

Première partie - Protection de la biodiversité

MPRO - RODD

ENCADREMENT :

Amélie Lambert

Cnam - Laboratoire Cedric

SOURCES : Alain Billionnet

Introduction

Deux grands sujets environnementaux préoccupent actuellement la communauté internationale : l'érosion de la biodiversité et le changement climatique, deux phénomènes liés au développement des sociétés humaines. Ces sujets ont pris beaucoup d'importance depuis la Conférence des Nations Unies sur l'Environnement et le Développement ou Sommet de la Terre de Rio de Janeiro en 1992. Nous nous intéressons ici au premier sujet. Il est actuellement admis que la biodiversité - des espèces, génétique et des écosystèmes - rend des services très importants aux sociétés humaines et que sa préservation est indispensable. L'assemblée générale des Nations Unies a d'ailleurs approuvé, fin 2010, la création de la Plate-forme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques, (IPBES, International science-policy Platform on Biodiversity and Ecosystem Services). Nous allons étudier quelques exemples illustrant l'aide que peut apporter la programmation mathématique aux décideurs en matière de protection de la biodiversité. Ces exemples concernent la sélection de réserves naturelles, la maîtrise des effets néfastes engendrés par la fragmentation des paysages, le maintien de la diversité génétique et l'exploitation écologique des forêts.



Projet 1 — *Sélection de réserves naturelles*

De nombreux pays se sont engagés à stopper la perte de biodiversité dans un avenir proche et ont adopté pour cela différentes stratégies dont la protection d'aires terrestres et maritimes. Ces aires protégées - ou réserves naturelles - contribuent de façon décisive au maintien de la biodiversité car elles visent directement la protection des éléments dont le risque de disparition est le plus fort. Ces éléments concernent la faune, la flore, les roches, minéraux et fossiles, ou encore de grands sites géomorphologiques. L'objectif est d'assurer à chaque espèce ou site menacé un espace où son avenir est garanti. Ainsi, de nombreux programmes gouvernementaux et non gouvernementaux cherchent à restaurer et protéger des habitats dans le but de préserver des espèces. Lors de la 10ème conférence de la Convention sur la diversité biologique qui s'est tenue à Nagoya en octobre 2010 un plan de protection de la biodiversité pour 2020 a été adopté. Ce plan comporte 20 objectifs dont la restauration d'habitats dégradés et la définition d'aires protégées (terrestres, marines et côtières). Commentant ce plan, le président de l'organisation environnementale Conservatoire International, souligne que le problème n'est pas seulement quantitatif mais aussi qualitatif et qu'il faut donc protéger les zones les plus riches et les plus importantes en termes de biodiversité. Les ressources disponibles pour cette protection étant évidemment limitées, il est important de les utiliser de façon efficace.



Problème de base et variantes

On s'intéresse à un ensemble d'espèces à protéger, $E = \{e_1, e_2, \dots, e_p\}$, vivant sur un ensemble de parcelles - ou sites -, $S = \{s_1, s_2, \dots, s_n\}$, répartis sur un territoire. Pour chaque parcelle, on connaît toutes les espèces qui y vivent et l'on suppose que ces espèces survivront dans cette parcelle si celle-ci est protégée. Le problème de base consiste à déterminer le plus petit sous-ensemble de parcelles permettant de protéger toutes les espèces considérées. Associons à chaque parcelle s_i une variable booléenne x_i qui vaut 1 si et seulement si cette parcelle est retenue pour être protégée. Posons $P = \{1, \dots, p\}$ et $N = \{1, \dots, n\}$. La détermination d'une réserve optimale, c'est-à-dire d'un sous-ensemble optimal de parcelles, peut se formuler par le programme mathématique :

$$(P1) \begin{cases} \min \sum_{i \in N} x_i \\ \text{s.t.} \\ \sum_{i \in S_k} x_i \geq 1 & k \in P \\ x \in \{0, 1\}^n \end{cases}$$

où S_k désigne l'ensemble des indices des parcelles permettant de protéger l'espèce e_k . La fonction objectif vise à minimiser le nombre de parcelles retenues et les p contraintes d'inégalité expriment que, pour chaque espèce considérée, au moins une parcelle la contenant doit être retenue. Le programme linéaire en variables binaires (P1) est connu en recherche opérationnelle sous le nom de **problème de recouvrement (Set Covering Problem)**. Il existe de nombreuses variantes de (P1). On peut chercher à minimiser l'aire totale de la réserve ou encore son coût total lorsqu'un coût est attribué à chaque parcelle, plutôt que le nombre de parcelles. La définition d'une réserve optimale peut aussi être abordée en cherchant à déterminer un sous-ensemble de parcelles qui maximise le nombre d'espèces protégées. Dans ce cas, le nombre de parcelles de la réserve ou son aire ou encore son coût est limité. On obtient alors le programme

$$(P2) \left\{ \begin{array}{l} \max \sum_{k \in P} y_k \\ \text{s.t.} \\ y_k \leq \sum_{i \in S_k} x_i \quad k \in P \\ \sum_{i \in N} a_i x_i \leq B \\ x \in \{0, 1\}^n \\ y \in \{0, 1\}^n \end{array} \right.$$

Dans une solution optimale de (P2), la variable booléenne y_k prend la valeur 1 ssi l'espèce e_k est présente dans au moins une des parcelles de la réserve. Les contraintes de (P1) et (P2) peuvent être ajustées de façon à imposer que chaque espèce soit représentée dans au moins r parcelles plutôt que dans une seule. Dans (P1), le second membre des contraintes devient r et dans (P2) les contraintes $y_k \leq \sum_{i \in S_k} x_i$ deviennent $ry_k \leq \sum_{i \in S_k} x_i$. Polasky et al. (2001)¹ ont utilisé ce modèle pour étudier la protection de 415 espèces de vertébrés terrestres réparties dans 289 parcelles de l'état de l'Oregon aux États-Unis.

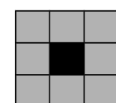
Réserves naturelles avec zone centrale et zone tampon

On considère deux types d'espèces à protéger, des **espèces en danger** et des **espèces communes**, et l'on distingue deux types de zones protégées, des **zones centrales** et des **zones tampons** qui protègent la zone centrale des nuisances entraînées par les activités humaines extérieures à la réserve. Pour être protégée (avec une certaine probabilité), une **espèce en danger** doit être présente dans une **zone centrale** et une **espèce commune** doit être présente **soit dans une zone centrale soit dans une zone tampon**. On considère qu'une parcelle sélectionnée fait partie d'une zone centrale si toutes les parcelles qui l'entourent sont elles aussi sélectionnées soit dans la zone centrale soit dans la zone tampon (le décideur peut définir une zone tampon plus ou moins importante). On connaît la **probabilité p_{ki}** que l'**espèce e_k** , présente dans la parcelle s_i , **survive dans cette parcelle si celle-ci est protégée**. On suppose que ces probabilités sont **indépendantes** et que $0 \leq p_{ki} \leq 1$. Le problème consiste à déterminer un sous-ensemble de parcelles de **coût minimal** et tel que, pour chaque espèce e_k , la **probabilité de présence** dans la réserve, c'est-à-dire dans ce sous-ensemble de parcelles, soit **supérieure ou égale** à une valeur donnée α_k .

Travail demandé

- i) Modéliser le problème de la détermination d'une réserve optimale avec zone centrale et zone tampon par un programme linéaire en variables 0-1.
- ii) Résoudre l'instance présentée ci-dessous. Elle comporte 3 espèces rares numérotées de 1 à 3 et 3 espèces communes numérotées de 4 à 6. On étudiera les cas suivants :
 - $\alpha_k = 0.5$ ($k = 1, \dots, 6$),
 - $\alpha_k = 0.9$ ($k = 1, \dots, 3$), et $\alpha_k = 0.5$ ($k = 4, \dots, 6$),
 - $\alpha_k = 0.5$ ($k = 1, \dots, 3$), et $\alpha_k = 0.9$ ($k = 4, \dots, 6$),
 - $\alpha_k = 0.8$ ($k = 1, \dots, 3$), et $\alpha_k = 0.6$ ($k = 4, \dots, 6$).

La zone géographique considérée est représentée par une matrice de 10×10 parcelles carrées et une parcelle retenue appartient à une zone centrale de la réserve ssi les 8 parcelles qui l'entourent font partie de la réserve.



1. Polasky, S., Camm, J.D., Garber-Yonts, B. (2001). Selecting biological reserves cost-effectively : an application to terrestrial vertebrate conservation in Oregon. Land Economics, 77, 68-78.

- iii) Pour chacune des 4 instances, donner le temps de calcul, le nombre de noeuds développés dans l'arbre de recherche, le coût de la solution et la probabilité de survie de chaque espèce dans la réserve.
- iv) Étudier le comportement du programme linéaire en variables 0-1 en fonction de la taille des instances. Pour cela, on engendrera des jeux d'essai aléatoirement.
- v) Écrire le modèle légèrement différent qui correspond à la détermination, sous une contrainte budgétaire, d'un sous-ensemble de parcelles tel que l'espérance mathématique du nombre d'espèces qui survivront dans la réserve formée soit maximale. Peut-on formuler ce problème par un programme linéaire en variables mixtes ?

Description des instances et solutions

Instance

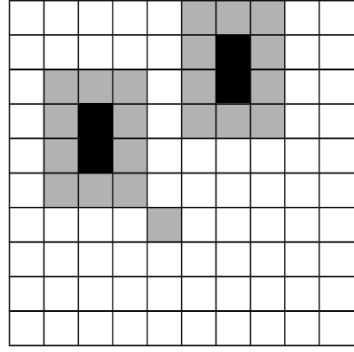
6	6	6	4	4	4	4	8	8	8
6	6	6	4	4	4	4	8	8	8
6	6	6	4	4	4	4	8	8	8
5	5	5	3	3	3	3	7	7	7
5	5	5	3	3	3	3	7	7	7
5	5	5	3	3	3	3	7	7	7
5	5	5	3	3	3	3	7	7	7
4	4	4	6	6	6	6	5	5	5
4	4	4	6	6	6	6	5	5	5
4	4	4	6	6	6	6	5	5	5

Coût de protection de chaque parcelle en unités de coût

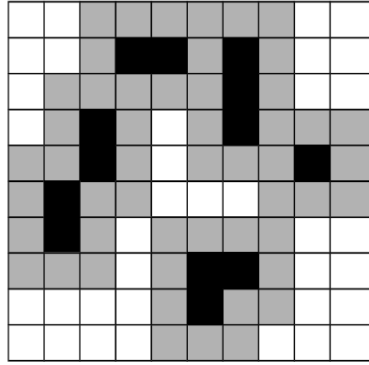
	4:0.3	6:0.4	4:0.3 6:0.4		6:0.5	6:0.3	6:0.3		5:0.4
5:0.3			3:0.2	3:0.3		2:0.2 3:0.4		6:0.2	5:0.3
	4:0.4			5:0.5 6:0.4		2:0.4 3:0.4		6:0.4	6:0.2
		1:0.4 4:0.4				2:0.2 3:0.5		6:0.3	
4:0.3		1:0.3 4:0.2		4:0.2 6:0.4				2:0.4 3:0.2	2:0.3 3:0.4
	1:0.4 4:0.2	4:0.4 5:0.4			5:0.2	5:0.2			
	2:0.5			4:0.4 5:0.4				4:0.3 5:0.5	
6:0.5					1:0.3	1:0.2 4:0.2		4:0.5 5:0.4	
6:0.2	3:0.3 5:0.5	5:0.2	5:0.4	1:0.2 5:0.4	1:0.4 2:0.2	2:0.2 4:0.4			
		6:0.2							

Probabilités de survie de chaque espèce dans chaque parcelle (**espèce** :probabilité)

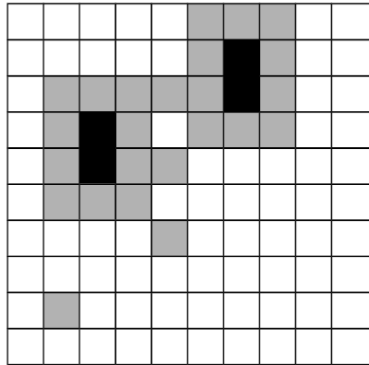
Solutions



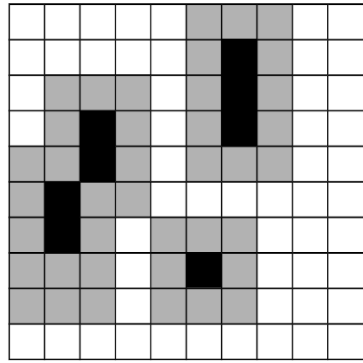
$\alpha_k = 0.5$ ($k = 1, \dots, 6$), coût = 119
 Probabilités de survie : (0.58, 0.52, 0.64, 0.92, 0.64, 0.76)



$\alpha_k = 0.9$ ($k = 1, \dots, 3$), $\alpha_k = 0.5$ ($k = 4, \dots, 6$), coût = 327
 Probabilités de survie : (0.92, 0.91, 0.92, 0.98, 0.89, 0.98)



$\alpha_k = 0.5$ ($k = 1, \dots, 3$), $\alpha_k = 0.9$ ($k = 4, \dots, 6$), coût = 130
 Probabilités de survie : (0.58, 0.52, 0.64, 0.93, 0.91, 0.91)



$\alpha_k = 0.8$ ($k = 1, \dots, 3$), $\alpha_k = 0.6$ ($k = 4, \dots, 6$), coût = 211
 Probabilités de survie : (0.82, 0.81, 0.82, 0.97, 0.78, 0.88)

Projet 2 — *Maîtrise des effets néfastes engendrés par la fragmentation du paysage*

La fragmentation des espaces est devenue une question centrale en écologie puisqu'elle est reconnue comme étant la principale source de perte de biodiversité dans les pays industrialisés. Elle entraîne une diminution des habitats et leur éclatement en fragments disjoints, affectant ainsi les populations de nombreuses espèces : extinction des espèces rares, augmentation des espèces communes et sédentaires, mortalité lors des déplacements entre fragments, érosion génétique, etc. Une part importante de cette fragmentation est provoquée par les infrastructures routières et ferroviaires mais elle est également due au développement urbain, à l'agriculture et à l'exploitation des forêts.



Les massifs boisés ne sont plus physiquement connectés. De nombreux invertébrés seront isolés dans les boisements. Sangliers et cervidés peuvent encore facilement circuler de l'un à l'autre.

Minimisation de la fragmentation

La fragmentation du paysage, c'est-à-dire l'éclatement de grandes superficies en petites parcelles doit être évaluée et contrôlée puisqu'elle influence de nombreux processus écologiques. Elle peut être mesurée par de multiples indicateurs. Ici nous considérons, à titre d'exemple, un indicateur courants, *la distance moyenne au plus proche voisin* (DMPPV). Cet indicateur fait partie de ceux qui sont utilisés dans FRAGSTATS (Marks et McGarigal, 1994²), un logiciel bien connu en écologie du paysage. Les problèmes d'optimisation associés à cette notion de fragmentation consistent, par exemple, à sélectionner, sous certaines contraintes, un sous-ensemble de parcelles optimal vis-à-vis des indicateurs retenus. Les parcelles sélectionnées seront, par exemple, conservées ou restaurées en tant que parcelles habitats et les autres seront affectés à d'autres usages (développement urbain, zones agricoles, coupes forestières, etc.). Voyons maintenant précisément la mesure de fragmentation que nous considérons ici, pour un ensemble de parcelles $S = \{s_1, s_2, \dots, s_n\}$:

$$\text{DMPPV} = \frac{1}{n} \sum_{i=1}^n \min \{d_{ij} : j = 1, \dots, n; j \neq i\}$$

où d_{ij} est la distance entre les deux parcelles s_i et s_j .

De nombreuses études ont montré que les mouvements des individus d'une même espèce entre différentes parcelles affectent la dynamique de la population et il est clair que ces mouvements sont plus difficiles si les valeurs de DMPPV sont élevées. Minimiser la valeur de cet indicateur favorise le regroupement de parcelles qui ne sont pas trop éloignées les unes des autres. En revanche, la distance entre les groupes de parcelles n'est pas contrôlée. Nous considérons ici le problème suivant associé à un ensemble de parcelles $S = \{s_1, s_2, \dots, s_n\}$ d'aire totale A : sélectionner un sous-ensemble de S , d'aire totale comprise entre A_{\min} et A_{\max} , de coût total inférieur ou égal à une certaine valeur B et qui minimise DMPPV.

2. Marks, B.J., McGarigal, K. (1994). Fragstats : Spatial pattern analysis program for quantifying landscape structure. Technical report, Forest Science Department, Oregon State University.

Optimisation combinatoire fractionnaire

Un problème d'optimisation combinatoire fractionnaire s'exprime sous la forme P :

$$(P) \begin{cases} \max \frac{f(x)}{g(x)} \\ \text{s.t.} \\ x \in X \subseteq [0, 1]^n \end{cases}$$

$f(x)$ et $g(x)$ sont deux fonctions de $[0, 1]^n$ dans \mathbb{R} ; X est un sous-ensemble de $[0, 1]^n$. On suppose que $g(x) > 0$ pour tout x appartenant à X et, également, qu'il existe un x de X tel que $f(x) > 0$.

Considérons le problème

$$(Q) \begin{cases} \min \lambda \\ \text{s.t.} \\ f(x) - \lambda g(x) \leq 0 \quad \forall x \in X \subseteq [0, 1]^n \\ \lambda \in \mathbb{R} \end{cases}$$

Un couple (λ^*, x^*) appartenant à $\mathbb{R} \times X$ est une solution optimale de (Q) si et seulement si

$$\forall x \in X, f(x) - \lambda^* g(x) \leq 0 = f(x^*) - \lambda^* g(x^*)$$

Cette condition est équivalente à

$$\forall x \in X, \frac{f(x)}{g(x)} \leq \lambda^* = \frac{f(x^*)}{g(x^*)}$$

Cette dernière condition signifie que λ^* est la valeur optimale du problème (P) et que x^* est une solution optimale de ce problème. Soit λ un paramètre appartenant à \mathbb{R} . Beaucoup de méthodes itératives visant à résoudre le problème (P) résolvent une suite d'instances du problème paramétrique (P_λ) associé à (P) :

$$(P_\lambda) \begin{cases} \max f(x) - \lambda g(x) \\ \text{s.t.} \\ x \in X \subseteq [0, 1]^n \end{cases}$$

Notons $v(\lambda)$ la valeur optimale de (P_λ) et x_λ^* une solution optimale de ce problème. On peut démontrer que $v(\lambda) = 0$ si et seulement si (λ, x_λ^*) est une solution optimale de (Q) , c'est-à-dire si et seulement si λ est la valeur optimale de (P) et x_λ^* une solution optimale de ce problème.

On obtient ainsi une autre formulation du problème (P) :

$$\text{Trouver } \lambda \in \mathbb{R} \text{ tel que } v(\lambda) = 0, \text{ où, } v(\lambda) = \max \left\{ f(x) - \lambda g(x) : x \in X \subseteq [0, 1]^n \right\}$$

A partir de cette formulation, on va pouvoir construire des algorithmes, pour résoudre le problème (P) , fondés sur des méthodes classiques permettant de déterminer la racine d'une fonction (méthode de Newton). Les propriétés suivantes de la fonction v découlent du fait que v correspond au maximum d'un nombre fini de fonctions linéaires décroissantes :

- (i) la fonction v est continue sur $] -\infty, +\infty[$ et strictement décroissante de $-\infty$ à $+\infty$.
- (ii) $v(0) > 0$, c'est une conséquence du fait qu'il existe un x de X tel que $f(x) > 0$.
- (iii) la fonction v a une et une seule racine λ^* et $\lambda^* > 0$.

- (iv) si $\lambda_1 < \lambda_2 < \dots < \lambda_q$ désignent toutes les valeurs de λ pour lesquelles deux droites de $\left\{ f(x) - \lambda g(x) : x \in X \subseteq [0, 1]^n \right\}$ se coupent, alors la fonction v est linéaire sur les intervalles $] -\infty, \lambda_1], [\lambda_i, +\lambda_{i+1}], (i = 1, 2, \dots, q-1)$ et $[\lambda_q, +\infty[$.
- (v) la fonction v est une fonction convexe.

Algorithme de Dinkelbach

1. $\lambda = \lambda_0$
2. calculer $v(\lambda) = \max \left\{ f(x) - \lambda g(x) : x \in X \subseteq [0, 1]^n \right\}$
3. soit x_λ tel que $v(\lambda) = f(x_\lambda) - \lambda g(x_\lambda)$
4. **si** $v(\lambda) > 0$ alors $\lambda \leftarrow \frac{f(x_\lambda)}{g(x_\lambda)}$ et **retour en 2**
sinon x_λ est une solution optimale de (P)

Dans le cas particulier où les deux fonctions $f(x)$ et $g(x)$ sont linéaires, c'est-à-dire $f(x) = a_1x_1 + a_2x_2 + \dots + a_nx_n$ et $g(x) = b_1x_1 + b_2x_2 + \dots + b_nx_n$ (problème d'optimisation combinatoire fractionnaire linéaire) on a le théorème suivant :

Théorème³

Le nombre d'itération de l'algorithme de Dinkelbach appliqué à (P) lorsque $f(x)$ et $g(x)$ sont linéaires est en $\mathcal{O}(n^2 \log(n)^2)$ ou $\mathcal{O}(n^3 \log(n))$.

Il est clair que, dans ce cas, le problème d'optimisation de l'étape 2 de l'algorithme consiste à résoudre un programme linéaire en nombres entiers.

Travail demandé

- i) Modéliser le problème de la sélection d'un ensemble de parcelles d'aire totale comprise entre A_{min} et A_{max} , de coût total inférieur ou égal à B et qui minimise DMPPV.
- ii) Écrire le modèle sous la forme d'un programme d'optimisation combinatoire fractionnaire, où les fractions sont des ratios de deux fonctions linéaires.
- iii) Résoudre l'instance présentée ci-dessous par l'algorithme de Dinkelbach pour les contraintes d'aire et de budget suivantes :
 - $A_{min} = 30, A_{max} = 35, B = 920$,
 - $A_{min} = 20, A_{max} = 21, B = 520$,
 - $A_{min} = 70, A_{max} = 75, B = 3500$.
- iv) Pour chacune des 3 instances, donner le temps de calcul total, le nombre total de noeuds développés dans l'arbre de recherche, le nombre d'itérations de l'algorithme de Dinkelbach, la valeur de la solution (valeur de DMPPV) et les parcelles retenues.
- v) Étudier le comportement de la méthode proposée en fonction de la taille des instances. Pour cela, on engendrera aléatoirement des jeux d'essai.

Description des instances et solutions

Instance

1	7	3	10	10	2	8	6	4	5	5
2	7	7	10	5	2	8	6	3	9	9
3	7	3	4	6	3	2	4	9	7	8
4	6	2	7	6	4	7	5	10	7	8
5	2	4	3	4	9	6	4	9	8	4
6	7	5	2	9	8	9	5	6	10	10
7	5	2	3	7	9	9	4	9	6	3
8	5	2	9	4	2	8	6	9	3	4
9	9	6	5	4	5	6	8	9	6	6
10	8	8	7	7	3	5	8	3	9	9

Zone représentée par une matrice de 10×10 parcelles carrées de 1 km de côté ; coût de sélection de de chaque parcelle en dizaine d'unités de coût

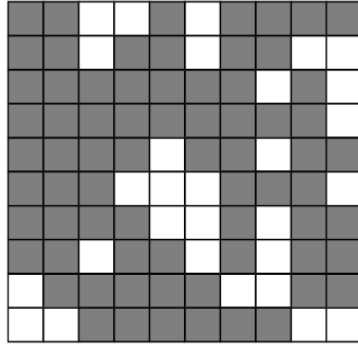
Solutions

	1	2	3	4	5	6	7	8	9	10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Solution de la première instance : $A_{min} = 30$, $A_{max} = 35$, $B = 920$:
Nb itérations = 2, nb noeuds = 0, nb parcelles sélectionnées = 30, DMPPV = 1.15501

	1	2	3	4	5	6	7	8	9	10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

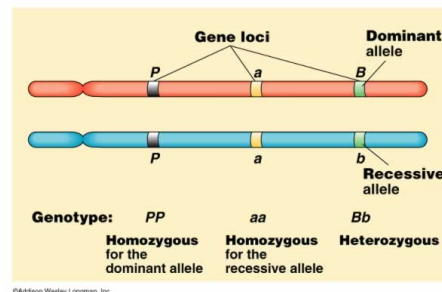
Solution de la deuxième instance : $A_{min} = 20$, $A_{max} = 21$, $B = 520$:
Nbitérations = 2, nb noeuds = 0, nbparcelles sélectionnées = 20, DMPPV = 1.27394



Solution de la troisième instance : $A_{min} = 70$, $A_{max} = 75$, $B = 3500$:
Nb itérations = 2, nb noeuds = 0, nb parcelles sélectionnées = 71, DMPPV = 1

Projet 3 — *Protection de la diversité génétique*

Le gène est l'unité de l'hérédité génétique. Il est caractérisé par son emplacement précis - ou locus - sur un chromosome. Un exemple bien connu est le gène de la couleur des yeux ou du groupe sanguin chez les êtres humains. Un gène correspond également à une séquence de nucléotides, c'est-à-dire à des fragments unitaires d'ADN. Un allèle est une version d'un gène. Ainsi la couleur des yeux chez les humains dépend des allèles portés par l'individu sur le gène concerné. La diversité génétique correspond à l'existence d'allèles différents d'un même gène. Dans les cellules diploïdes (n paires de chromosomes), c'est-à-dire dans les cellules des êtres humains et de la plupart des animaux, chaque individu a deux allèles - correspondant à une paire de chromosomes - pour chaque gène et un individu reçoit, au hasard, un des deux allèles de chacun de ses parents, correspondant à un des deux chromosomes de ses parents. Ces allèles peuvent être identiques ou non. Prenons l'exemple du groupe sanguin avec, par exemple, la situation suivante : un individu A/B a hérité au hasard l'allèle A d'un des parents qui était A/O et l'allèle B de l'autre parent qui était B/O . Un individu dont un gène présente des allèles identiques sur les deux chromosomes est dit homozygote pour ce gène. Dans le cas contraire, on dit que l'individu est hétérozygote pour ce gène. Dans ce dernier cas, il y a deux possibilités pour le phénotype (état d'un caractère observable) : un des allèles est dominant et l'autre est récessif ou ce n'est pas le cas. Si plus de deux allèles correspondent à un même gène, le gène est polyallélique. La diversité génétique est un des facteurs permettant aux espèces de s'adapter à l'évolution de leur environnement tout particulièrement dans le cadre du changement climatique et des changements globaux. De nombreux articles publiés dans la littérature de la conservation biologique et consacrés aux programmes de conservation des espèces traitent de problèmes d'optimisation de la diversité génétique. Un type de problèmes concerne la gestion de la reproduction de populations en captivité, de façon à maintenir le plus possible la diversité génétique. Nous présentons ci-dessous un problème de ce type.



Maintien de la diversité allélique dans la descendance d'une population

Nous présentons ici un problème classique rencontré dans les programmes de conservation génétique. Il consiste à déterminer la contribution (nombre de progénitures) optimale de chaque individu de façon à minimiser la perte d'allèles dans la population engendrée. On considère une population formée de P individus, I_1, I_2, \dots, I_P , appartenant à une espèce donnée et un ensemble de M gènes polyalléliques à prendre en compte, g_1, g_2, \dots, g_M . Pour chaque gène g_i , les allèles possibles sont $a_{i_1}, a_{i_2}, \dots, a_{i_{t(i)}}$. On se place dans le cas de cellules diploïdes (être humain et la plupart des animaux) où, pour chaque gène, chaque individu a deux allèles, chaque allèle correspondant à un chromosome de la paire considérée. Ces allèles peuvent être identiques ou non. Chaque individu reçoit au hasard un des deux allèles de chacun de ses parents (voir la Figure 1 ci-dessous). On cherche à déterminer le nombre optimal de progénitures de chaque individu de façon à minimiser l'espérance mathématique du nombre d'allèles disparus dans la génération suivante tout en conservant une population de taille constante.

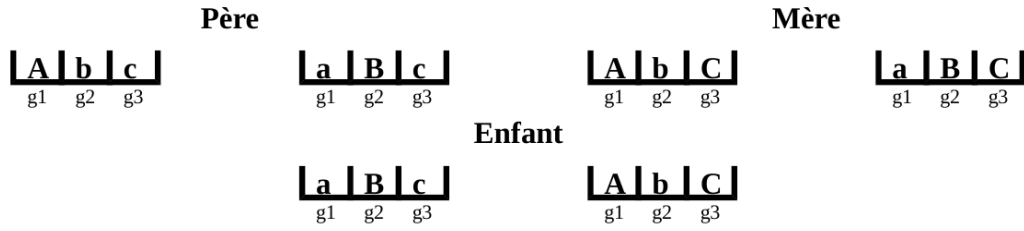
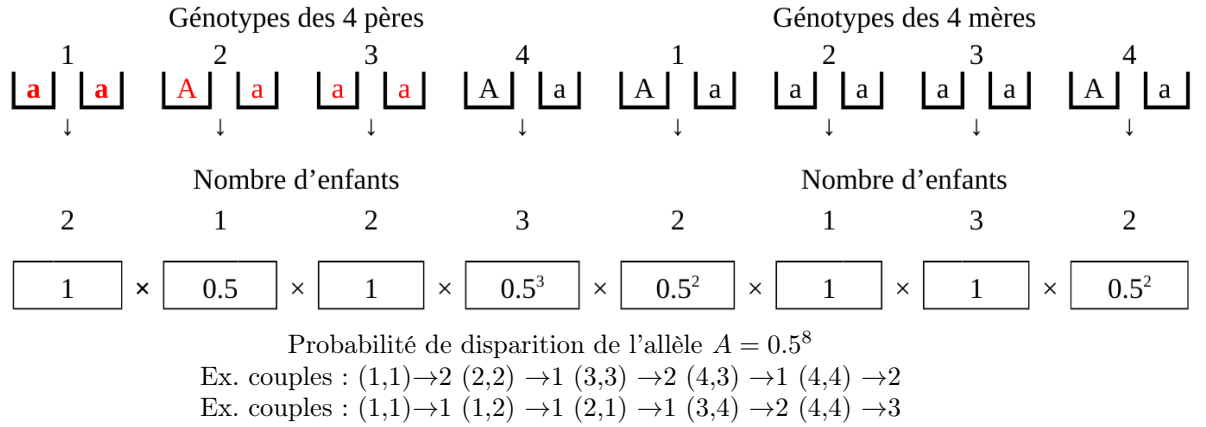


FIGURE 1 – Transmission de 3 gènes bi-alléliques $g_1(A/a)$, $g_2(B/b)$ et $g_3(C/c)$ sur une paire de chromosomes.

Exemple de calcul de la probabilité de disparition d'un allèle donné (sur un locus donné)



Approximation de la fonction logarithmique par une fonction linéaire par morceaux

Considérons la contrainte :

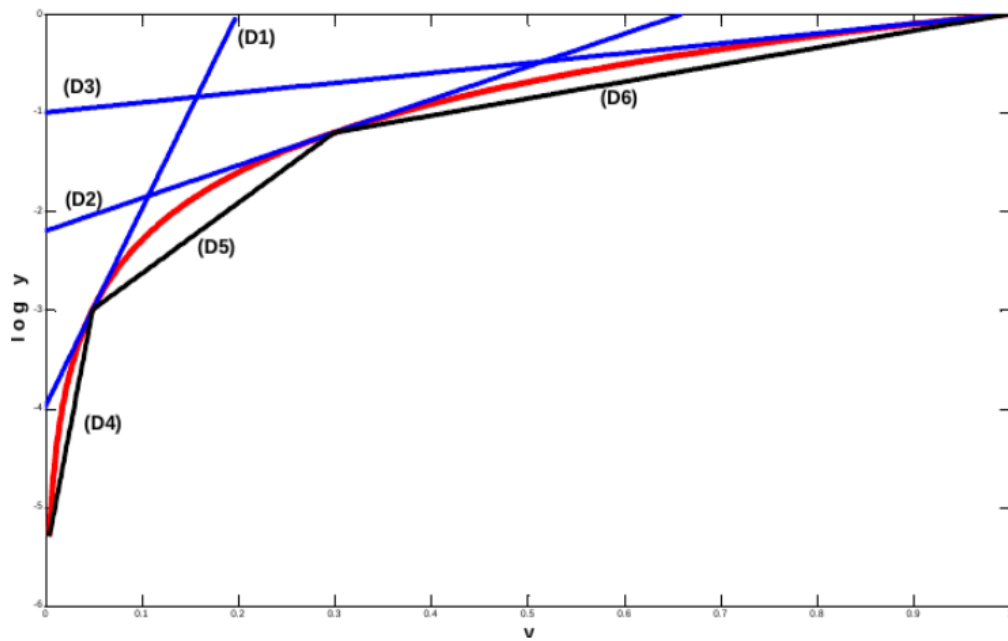
$$\log(y) \geq f(x_1, x_2, \dots, x_n) \quad (1)$$

où $0 < y \leq 1$ et $x \in \mathbb{R}^n$.

L'ensemble de contraintes (C) constitue une **relaxation** de la contrainte (1) :

$$(C) : \log(\theta_r) + \frac{1}{\theta_r}(y - \theta_r) \geq f(x_1, x_2, \dots, x_n) \quad (r \in H)$$

où $H = \{1, 2, \dots, h\}$ et θ est un vecteur donné de \mathbb{R}^h tel que $0 < \theta_1 < \theta_2 < \dots < \theta_h = 1$.



Deux approximations de $\log(y)$ ($0 < y \leq 1$), par une fonction linéaire par morceaux.

Les droites (D1), (D2) et (D3) sont les contraintes C .

Les droites (D4), (D5) et (D6) une approximation inférieure.

Travail demandé

- i) Modéliser le problème de la détermination du nombre optimal de progénitures de chaque individu de façon à minimiser l'espérance mathématique du nombre d'allèles disparus dans la génération suivante tout en conservant une population de taille constante.
- ii) Associer au modèle précédent un programme linéaire en variables mixtes permettant de trouver simultanément une solution admissible du problème et une borne inférieure de la valeur de la solution optimale. Idée : approximer la fonction logarithmique par une fonction linéaire par morceaux.
- iii) Résoudre l'instance présentée ci-dessous en limitant le nombre de progénitures de chaque individu à 3 puis à 2.
- iv) Pour chacune des 2 instances, donner le temps de calcul, le nombre de noeuds développés dans l'arbre de recherche, la probabilité de disparition de chaque allèle, l'espérance du nombre d'allèles disparus et la valeur de la borne inférieure.
- v) Étudier le comportement du programme linéaire en fonction de la taille des instances et du nombre de morceaux considérés dans la fonction linéaire par morceaux approxinant la fonction logarithmique, en engendrant aléatoirement des jeux d'essai.

Description des instances et solutions

Instance

Nombre d'individus dans la population : 8 (4 mâles / 4 femelles)

Nombre de paires de chromosomes (par individu) considérées : 1

Nombre de locus considérés par chromosome : 5

Nombre d'allèles par gène : 2

M Â L E S	1	<table><tr><td>a</td><td>B</td><td>c</td><td>d</td><td>E</td></tr></table>	a	B	c	d	E	<table><tr><td>A</td><td>B</td><td>C</td><td>D</td><td>e</td></tr></table>	A	B	C	D	e
	a	B	c	d	E								
	A	B	C	D	e								
	2	<table><tr><td>a</td><td>b</td><td>C</td><td>D</td><td>E</td></tr></table>	a	b	C	D	E	<table><tr><td>a</td><td>B</td><td>c</td><td>d</td><td>E</td></tr></table>	a	B	c	d	E
a	b	C	D	E									
a	B	c	d	E									
3	<table><tr><td>a</td><td>b</td><td>c</td><td>D</td><td>E</td></tr></table>	a	b	c	D	E	<table><tr><td>a</td><td>B</td><td>C</td><td>d</td><td>e</td></tr></table>	a	B	C	d	e	
a	b	c	D	E									
a	B	C	d	e									
4	<table><tr><td>a</td><td>B</td><td>C</td><td>d</td><td>e</td></tr></table>	a	B	C	d	e	<table><tr><td>a</td><td>B</td><td>c</td><td>d</td><td>e</td></tr></table>	a	B	c	d	e	
a	B	C	d	e									
a	B	c	d	e									
F E M E L L E S	1	<table><tr><td>a</td><td>B</td><td>C</td><td>d</td><td>E</td></tr></table>	a	B	C	d	E	<table><tr><td>A</td><td>B</td><td>C</td><td>d</td><td>e</td></tr></table>	A	B	C	d	e
	a	B	C	d	E								
	A	B	C	d	e								
	2	<table><tr><td>A</td><td>B</td><td>c</td><td>d</td><td>e</td></tr></table>	A	B	c	d	e	<table><tr><td>a</td><td>B</td><td>c</td><td>D</td><td>E</td></tr></table>	a	B	c	D	E
A	B	c	d	e									
a	B	c	D	E									
3	<table><tr><td>A</td><td>B</td><td>c</td><td>D</td><td>E</td></tr></table>	A	B	c	D	E	<table><tr><td>A</td><td>B</td><td>c</td><td>D</td><td>E</td></tr></table>	A	B	c	D	E	
A	B	c	D	E									
A	B	c	D	E									
4	<table><tr><td>a</td><td>B</td><td>c</td><td>d</td><td>e</td></tr></table>	a	B	c	d	e	<table><tr><td>a</td><td>B</td><td>c</td><td>D</td><td>E</td></tr></table>	a	B	c	D	E	
a	B	c	d	e									
a	B	c	D	E									

Solutions

Solution optimale avec la contrainte $x_j \leq 3 \quad \forall j$							
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
2	3	3	0	3	1	1	3
Espérance du nombre d'allèles perdus 0.015869*							
Borne inférieure 0.015588*							
Probabilité de disparition des allèles A, a, B, C, c, D, d, E = 0 Probabilité de disparition de l'allèle b = 0.015625, de l'allèle e = 0.000244141							
Solution optimale avec la contrainte $x_j = 2 \quad \forall j$							
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
2	2	2	2	2	2	2	2
Espérance du nombre d'allèles perdus 0.062500*							
Borne inférieure 0.062433*							
Probabilité de disparition des allèles A, a, B, C, c, D, d, E, e = 0 Probabilité de disparition de l'allèle b = 0.062500 (=0.5⁴)							

Résultats obtenus avec $\theta_r = \theta_1^{\frac{h-r}{h-1}}$ où $\theta_1 = 0.001$ et $h = 50$.

Remarque : En prenant $\theta_r = \theta_1^{\frac{h-r}{h-1}}$ où $\theta_1 = 0.0001$ et $h = 50$, on obtient, pour l'instance où $x_j \leq 3$ pour tout j , une solution admissible de valeur 0.015625 et une borne inférieure de valeur 0.015621. La seule probabilité non nulle de disparition d'un allèle correspond à l'allèle b et est égale à 0.015625.

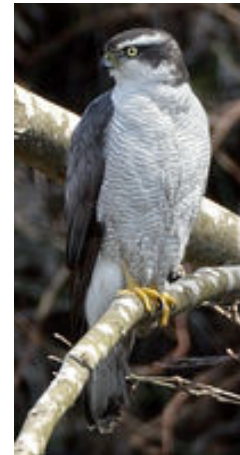
Projet 4 — *Exploitation durable de la forêt*

La gestion durable des forêts comporte un aspect économique, un aspect environnemental et un aspect bien être des populations. Depuis 1992, cette notion a été précisée par des conférences internationales et la France l'a reprise dans la loi d'orientation sur la forêt de 2001 : la gestion durable des forêts garantit leur diversité biologique, leur productivité, leur capacité de régénération, leur vitalité et leur capacité à satisfaire, actuellement et pour l'avenir, les fonctions économique, écologique et sociale pertinentes, aux niveaux local, national et international, sans causer de préjudices à d'autres écosystèmes. Les publications sur le sujet faisant intervenir la notion d'optimisation sont nombreuses. Nous étudions ci-dessous un problème précis d'exploitation durable de la forêt.



Exploitation de la forêt visant à protéger certaines espèces

Nous présentons dans ce paragraphe un modèle d'exploitation de la forêt visant à protéger le mieux possible certaines espèces. On considère un ensemble de parcelles forestières carrées et identiques représenté par une matrice $m \times n$ et deux types d'espèces e_1 et e_2 . On note l la longueur du côté des parcelles. L'habitat de l'espèce e_1 se situe principalement dans les parcelles coupées et l'habitat de l'espèce e_2 est constitué essentiellement des lisières entre parcelles coupées et parcelles non coupées. On considère la zone représentée par la matrice comme une zone non coupée et la zone extérieure à la matrice comme une zone coupée. La population attendue de l'espèce e_1 dans chaque parcelle s_{ij} coupée (resp. non coupée) est égale à t_{ij} (resp. 0) et la population attendue de l'espèce e_2 est égale à gl où g désigne la population attendue de l'espèce e_2 pour chaque kilomètre de lisière et L , la longueur totale de lisière compte tenu des coupes réalisées. Par exemple, l'autour des palombes aime cet habitat de lisières dans le voisinage duquel s'étendent des milieux ouverts où il peut chasser les petits mammifères vivant dans ce même habitat. Le problème consiste à déterminer les parcelles à couper et les parcelles à laisser en l'état de façon à maximiser la somme pondérée des populations des deux espèces.



Formulation par un programme linéaire en variables mixtes

Associons à chaque parcelle s_{ij} la variable booléenne x_{ij} qui vaut 1 ssi la parcelle est non coupée. On peut formuler le problème par le programme linéaire en variables 0-1 suivant :

$$(P1) \begin{cases} \max w_1 \sum_{(i,j) \in M \times N} t_{ij}(1 - x_{ij}) + w_2 gl \sum_{(i,j) \in M \times N} (4x_{ij} - d_{ij}) \\ \text{s.t.} \\ d_{ij} \geq \sum_{(k,l) \in A_{ij}} x_{kl} - |A_{ij}|(1 - x_{ij}) \\ d \in \mathbb{R}_+^{M \times N} \\ x \in \{0, 1\}^{M \times N} \end{cases} \quad (i, j) \in M \times N$$

où $M = 1, \dots, m$, $N = 1, \dots, n$, w_1 et w_2 sont les coefficients de pondération, l est la longueur du côté de chaque parcelle et A_{ij} désigne l'ensemble des couples (k, l) tels que la parcelle s_{kl} est adjacente à la parcelle s_{ij} .

Travail demandé

- i) Expliquer le programme linéaire en variables mixtes ($P1$).
- ii) Modéliser le problème par un programme quadratique en variables 0-1 noté ($P2$).
- iii) Proposer une linéarisation du programme ($P2$) dont la matrice des contraintes est totalement unimodulaire. On montrera que la matrice des contraintes de cette linéarisation est TU, en s'appuyant sur le fait que la matrice d'incidence sommets-arêtes d'un graphe biparti est TU⁴.
- iv) Résoudre les 2 instances présentées ci-dessous par les deux approches.
- v) Pour chacune des 2 approches, donner le temps de calcul, le nombre de noeuds développés dans l'arbre de recherche. Comparer les résultats obtenus et conclure.
- vi) Résoudre cette même instance par les deux approches dans le cas où l'on impose que le nombre de parcelles non coupées soit supérieur ou égal à 60. De façon générale, la matrice des contraintes reste-t-elle TU lorsqu'on ajoute cette contrainte sur le nombres de parcelles non coupées ?
- vii) Étudier le comportement des deux approches en fonction de la taille des instances. Pour cela, on engendrera aléatoirement des jeux d'essai.

4. Nemhauser, G., Wolsey, L.A. (1988). Combinatorial Optimization, Wiley-Interscience, New York.

Description des instances et solutions

Instances

84	68	97	98	64	89	82	71	74	76
87	83	98	75	60	90	78	67	92	94
84	68	70	81	67	61	73	92	86	90
79	62	86	79	73	84	76	98	84	90
62	72	66	72	92	80	71	91	87	70
85	77	63	93	90	94	76	81	99	98
76	63	66	84	94	93	72	92	79	65
76	63	92	69	60	88	79	93	66	73
92	82	77	72	77	81	89	95	80	80
88	89	83	86	69	78	91	64	94	92

$[t_{ij}]$

$m = n = 10, w_1 = 1, w_2 = 5, l = 3, g = 1.26157$

10	10	10	1	10
10	10	1	1	10
10	10	1	10	10
1	10	10	10	10
1	10	10	10	10

$[t_{ij}]$

$m = n = 5, w_1 = 2, w_2 = 1, l = 3, g = 1.26157$

Solutions

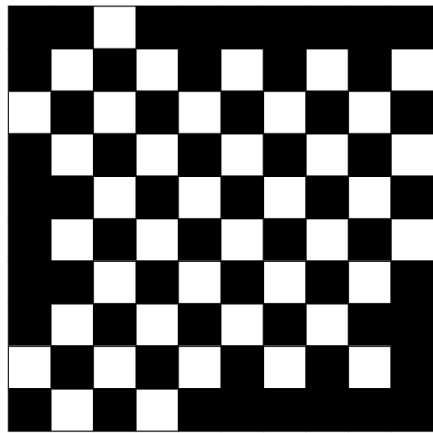
Effectif de l'espèce e_1 : 6630

Effectif de l'espèce e_2 : 317.92

Nombre de parcelles non coupées : 21

Nombre de côtés appartenant à la lisière : 84

Valeur de la fonction économique à l'optimum : 8219.6



Solution dans le cas où le nombre de parcelles non coupées est supérieur ou égal à 60.

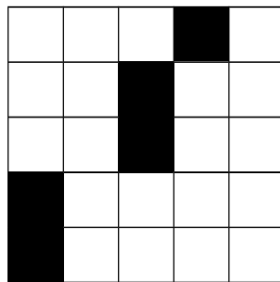
Effectif de l'espèce e_1 : 3272

Effectif de l'espèce e_2 : 696.39

Nombre de parcelles non coupées : 60

Nombre de côtés appartenant à la lisière : 184

Valeur de la fonction économique à l'optimum : 6753.93



Effectif de l'espèce e_1 : 191

Effectif de l'espèce e_2 : 60.56

Nombre de parcelles non coupées : 5

Nombre de côtés appartenant à la lisière : 16

Valeur de la fonction économique à l'optimum : 442.56