MILP pour la sélection d'une équipe optimale

Optimisation du score marqué par une équipe lors des interclubs de natation

Ce court manuscrit offre une approche structurée pour aborder le problème d'optimisation suivant : la conception d'une équipe maximisant le nombre de points tout en respectant les contraintes imposé par le règlement de la compétition. Il se concentre sur le développement d'un algorithme employant la programmation linéaire entière pour résoudre efficacement ce défi d'optimisation. Cette méthode est préférée à l'approche par force brute, qui se révèle trop consommatrice de ressources.

Guénolé CHÉROT
guenole.cherot@univ-rennes.fr

IUT de Rennes, Département GEII
2023 - 2024

Table des matières

1 Objectifs

L'objectif est de former l'équipe qui réalisera le plus de points au inter-club. Nous disposons des prévisions de performance pour chaque nageur sur chaque épreuve.

1.1 Épreuves

Les épreuves sont les suivantes [1] :

- Épreuves individuelles :
  - 50 NL
  - 100 NL
  - 200 NL
  - 50 Pap
  - 100 Pap
  - 50 Dos
  - 100 Dos
  - 50 Brasse
  - 100 Brasse
  - 100 4N
- Relais :
  - 10x50 NL
  - 4x50 4N

1.2 Contraintes

- Toute les nages doivent être nagées une et une seule fois.
- L'équipe est composée de 10 membres
- Tous les nageurs sélectionnés doivent nager une nage individuelle et participer au relais 10x50m NL.
- Le relais 4x50 4N doit être composé de deux hommes et deux femmes.

1.3 Score

Pour chaque course, un score est calculé en fonction de la table de cotation Extranat et du coefficient de rajeunissement [2] :

- Les points des épreuves individuelles seront calculés ainsi : Temps réalisé divisé par le coefficient de majoration en fonction de l'épreuve et de la catégorie et calcul des points à la table de cotation FFN.
- Les points des épreuves de relais seront calculés ainsi :
  - Pour une relayeuse, le coefficient spécifique est pris dans la table dames en relais.
  - Pour un relayeur, le coefficient est pris dans la table individuelle messieurs.
    - 4 x 50 m 4 nages : cotation FFN du relais messieurs 4 x 50 m 4N après correction par la moyenne des coefficients.
    - 10 x 50 m nage libre : cotation FFN du 10 x 50 m nage libre messieurs après correction par la moyenne des coefficients.

1. Règlement détaillé sur Extranat : lien.
2. Les deux sont disponible en csv sur le site extranat : lien

2 Type d'approche

2.1 Force brut

Le problème ne peut pas être résolue par des approches combinatoires à cause du nombre de nageurs dans le club et du nombre d'épreuve. Considérons la matrice de taille $n \times m$ où n représente le nombre de nageur et m le nombre de nage, chaque élément de la matrice peut prendre deux valeurs : 0 ou 1 selon si le nageur nage ou pas l'épreuve. Dans une première approximation, on peut considérer que le nombre de possibilités à tester correspond au nombre de matrice possible soit $2^{n \times m}$. Dans le cas du club "rennes natation" cela correspond à $2^{30 \times 10} \approx 10^{90}$ possibilités.

Une approche plus réaliste peut-être de choisir k nageur parmi n et ensuite faire des permutation dans les k nageur choisit pour trouver la meilleur combinaison de ces nageurs soit $k!$ permutation. Au total on a donc :

$$\binom{n}{k} \cdot k! = \frac{n!}{k!(n-k)!} \cdot k! = \frac{n!}{(n-k)!}$$

Dans notre cas on a 20 nageurs et 10 nageurs par équipe. Soit $7 \cdot 10^{11}$ possibilités. Les opérations étant simple, on peut espérer une fréquence de calcul de quelques MHz. Le calcul pourrait être résolu en $7 \cdot 10^{11}/10^{6} \approx 10^{6}s \approx 12$ jours.

2.2 Alternatives

La programmation linéraire entière permet de résoudre une grande variété de problèmes linéaires. Elle a l'avantage d'être très rapide à résoudre et est donc adaptée dans notre cas. La formulation proposée dans ce manuscrit s'inspire des travaux suivants : Khandelwal, S. N. (2019). Building Teams of Experts using Integer Linear Programming. University of Windsor. lien.

2.3 Formulation générale d'un problème MILP

Nous cherchons à formuler le problème sous la forme suivante :

$$\begin{aligned} &\min_x c^T x \\ \text{tel que } & b_l \leq Ax \leq b_u, \\ & l \leq x \leq u, \\ & x_i \in \mathbb{Z}, i \in \mathbb{Z} \end{aligned} \tag{1}$$

3 Formulation mathématique

3.1 Fonction objectif

3.1.1 Épreuves individuelles

On note $x_{i,j}$ la participation du nageur i à l'épreuve individuelle j. La variable binaire $x_{i,j}$ prend la valeur 1 en cas de participation, 0 sinon. On note n le nombre de nageur dans le club et m le nombre d'épreuves à disputer. On peut donc créer $\mathbf{X} \in \mathcal{M}_{n,m}$, la matrice de participation aux épreuves individuelles $\mathbf{X} \in \mathcal{M}_{n,m}$:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \cdots \\ \mathbf{x_n} \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \cdots & & & \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{pmatrix} \tag{2}$$

De la même manière, on peut définir la matrice de points $\mathbf{S} \in \mathcal{M}_{n,m}$ qui défini le nombre de points obtenu à chaque épreuve individuelle.

Le score des épreuves individuel de l'équipe se calcule de la manière suivante :

$$s_{indiv} = \mathbf{X} \circ \mathbf{S}$$

où $\circ$ est le produit matriciel de Hadamard, qui pour deux matrices de mêmes dimensions, associe une autre matrice, de même dimension, et où chaque coefficient est le produit terme à terme des deux matrices.

3.1.2 Relais

10x50 NL : Tous les participants seront inscrit au relais 10x50 NL. Nous pouvons donc créer une matrice de participation à la compétition notée $\mathbf{P} \in \mathcal{M}_{n,1}$ et défini par :

$$\mathbf{P} = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ p_n \end{pmatrix} \tag{3}$$

On note $\mathbf{T} \in \mathcal{M}_{n,1}$ la matrice de points au 50NL. Le score au relais 10x50 NL est donc donnée par $s_{NL} = \mathbf{P} \circ \mathbf{T}$.

4x50 quatre nages : Comme pour les épreuves individuelles, nous pouvons créer la matrice de participation au relais 4x50 quatre nages $\mathbf{Y} \in \mathcal{M}_{n,4}$ et les points associés $\mathbf{R} \in \mathcal{M}_{n,4}$. Le score au relais 4x50 quatre nages est donc donnée par $s_{4N} = \mathbf{Y} \circ \mathbf{R}$.

3.1.3 Score total

Pour finir le score total est donnée par

$$s_{tot} = \mathbf{X} \circ \mathbf{S} + \mathbf{P} \circ \mathbf{T} + \mathbf{Y} \circ \mathbf{R} \tag{4}$$

3.2 Contraintes

3.2.1 Chaque épreuve doit être nagée une fois

Pour chaque épreuve j, il ne doit y avoir qu'un seul participant :

$$\sum_{i}^{n} x_{i,j} = 1, \ \ \forall j \in [\![0; m]\!] \tag{5}$$

3.2.2 L'équipe est composée de 10 membres

La somme des participants doit être égale à 10 :

$$\sum_{i}^{n} p_i = 10 \tag{6}$$

3.2.3 Tous les nageurs sélectionnés doivent nager une nage individuelle et participer au relais 10x50m NL

Quatres cas de figure sont possibles :
- Participation à la compétition
  - Le nageur i a été sélectionné pour participer à la compétition : $p_i = 1$
  - Le nageur i n'a pas été sélectionné pour participer à la compétition : $p_i = 0$
- Participation aux épreuves individuelles
  - Le nageur i nage une épreuve individuel :

$$\exists! j \in [\![0; m]\!], \ x_{i,j} = 1 \implies \sum_{j}^{m} x_{i,j} = 1$$

  - Le nageur i ne nage pas d'épreuve individuel :

$$\forall j \in [\![0; m]\!], \ x_{i,j} = 0 \implies \sum_{j}^{m} x_{i,j} = 0$$

On peut donc construire la table ci-dessous :

$\sum_j^m x_{i,j}$	p_i	$\sum_j y_{i,} - p_i$	état valide
0	0	0	Oui
0	1	-1	Non
1	0	1	Non
1	1	0	Oui

On en déduit la contrainte suivante :

$$\left(\sum_{j}^{m} x_{i,j} \right) - p_i = 0, \ \ \forall i \in [\![0; n]\!] \tag{7}$$

3.2.4 Le relais 4x50 4N doit être composé de deux hommes et deux femmes.

Le nageur qui participe au relais 4x50 4N doit avoir été sélectionné à la compétition :
- Participation à la compétition :

— Le nageur i a été sélectionné pour participer à la compétition : $p_i = 1$
— Le nageur i n'a pas été sélectionné pour participer à la compétition : $p_i = 0$

— Participation au relais :
  — Le nageur i participe au relais :

$$\exists ! j \in [\![0; m]\!],\ y_{i,j} = 1 \implies \sum_j y_{i,j} = 1$$

  — Le nageur i ne participe pas au relais :

$$\forall j \in [\![0; m]\!],\ y_{i,j} = 0 \implies \sum_j y_{i,j} = 0$$

On peut donc construire la table ci-dessous :

$\sum_j y_{i,j}$	p_i	$2 \cdot \sum_j y_{i,} - p_i$	état valide
0	0	0	Oui
0	1	-1	Oui
1	0	2	Non
1	1	1	Oui

On en déduit la contrainte suivante :

$$2 \cdot \left(\sum_j^m y_{i,j} \right) - p_i \leq 1,\ \ \forall i \in [\![0; n]\!] \tag{8}$$

Nage nagé qu'une seule fois : Chaque nage j (pap, dos, brasse et crawl) ne peut être nagée qu'une fois par relais :

$$\sum_i Y_{i,j} = 1,\ \ \forall j \in [\![0; m]\!] \tag{9}$$

Parité de genre : Le relais doit contenir exactement deux hommes et deux femmes. On créer pour cela la une matrice de genre noté $\mathbf{G} \in \mathcal{M}_{n,1}$, ou $g_i = 1$ pour les nageuses et $g_i = -1$ pour les nageurs. La contrainte s'écrit alors :

$$\sum_j \sum_i Y_{i,j} g_i = 0 \tag{10}$$

4 Implémentation

Les fonctions coût et contraintes décrites dans la partie précédentes doit être exprimées sous la forme des produits matriciel. Cette section propose une implémentation.

4.1 Fonction coût

La fonction coût totale à maximiser à donnée par l'équation (4). Or pour pouvoir utiliser l'implémentation de scipy, il faut l'exprimer sous la forme :

$$\min c^T x$$

Une implémentation possible consiste à "applatir" les matrices **X**, **P** et **Y** et de les concaténer. On procède de la même manière pour la matrice des scores.

$$x = \begin{pmatrix} \mathbf{x_1}^T \\ \mathbf{x_2}^T \\ \dots \\ \mathbf{x_n}^T \\ \mathbf{P} \\ \mathbf{y_1}^T \\ \dots \\ \mathbf{y_n}^T \end{pmatrix}, \quad c = \begin{pmatrix} \mathbf{s_1}^T \\ \mathbf{s_2}^T \\ \dots \\ \mathbf{s_n}^T \\ \mathbf{T} \\ \mathbf{r_1}^T \\ \dots \\ \mathbf{r_n}^T \end{pmatrix} \tag{11}$$

4.2 Contraintes

4.2.1 Valeurs possibles pour x

Pour pouvoir utiliser l'implémentation de scipy, il faut contraindre l'espace de définition de x sous la forme :

$$l \leq x \leq u$$

. Dans notre cas c'est simple, x est une variable binaire. On a donc :

$$l = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} \tag{12}$$

4.2.2 Contraintes lié aux problèmes

Le code final doit implémenter les contraintes (5),(6),(7),(8),(9),(10). Elles doivent être exprimées sous la forme :

$$b_l \leq Ax \leq b_u$$

Chaque épreuve doit être nagée une fois

$$\forall j \in [\![0; m-1]\!], \quad \sum_i^n x_{i,j} = 1$$

$$\mathbf{A}^{[\mathbf{X},j]} = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ & & & \dots & & & \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ & & & \dots & & & \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}, \quad \mathbf{A}^{[\mathbf{P},j]} = \begin{pmatrix} 0 & \dots & 0 \end{pmatrix}, \tag{13}$$
$$\mathbf{A}^{[\mathbf{Y},j]} = \begin{pmatrix} 0 & \dots & 0 \\ & \dots & \\ 0 & \dots & 0 \end{pmatrix}$$

Pour la suite on posera toujours :

$$\forall j \in [\![0; m-1]\!], \quad \mathbf{A}^{[j]} = \begin{pmatrix} \mathbf{a}_1^{[\mathbf{X},j]} & \dots & \mathbf{a}_n^{[\mathbf{X},j]} & \mathbf{A}^{[\mathbf{P},j]} & \mathbf{a}_1^{[\mathbf{Y},j]} & \dots & \mathbf{a}_n^{[\mathbf{Y},j]} \end{pmatrix} \tag{14}$$

On a ici une contrainte d'égalité donc :

$$\forall j \in [\![0; m-1]\!], \;\; b_l^{[j]} = u_l^{[j]} = 1 \tag{15}$$

L'équipe est composée de 10 membres

$$\sum_i^n p_i = 10$$

$$\mathbf{A}^{[\mathbf{X}]} = \begin{pmatrix} 0 & ... & 0 \\ & ... & \\ 0 & ... & 0 \end{pmatrix}, \; \mathbf{A}^{[\mathbf{P}]} = \begin{pmatrix} 1 & ... & 1 \end{pmatrix}, \; \mathbf{A}^{[\mathbf{Y}]} = \begin{pmatrix} 0 & ... & 0 \\ & ... & \\ 0 & ... & 0 \end{pmatrix} \tag{16}$$

On a ici une contrainte d'égalité donc :

$$b_l = u_l = 10 \tag{17}$$

Tous les nageurs sélectionnées doivent nager une individuelle et participer au relais 10x50 NL

$$\forall i \in [\![0; n-1]\!], \quad \left(\sum_j^m x_{i,j} \right) - p_i = 0$$

$$\mathbf{A}^{[\mathbf{X},i]} = \begin{pmatrix} 0 & ... & 0 & 0 & 0 & ... & 0 \\ & & & ... & & & \\ 0 & ... & 0 & 0 & 0 & ... & 0 \\ 1 & ... & 1 & 1 & 1 & ... & 1 \\ 0 & ... & 0 & 0 & 0 & ... & 0 \\ & & & ... & & & \\ 0 & ... & 0 & 0 & 0 & ... & 0 \end{pmatrix}, \; \mathbf{A}^{[\mathbf{P},i]} = \begin{pmatrix} 0 & ... & 0 & 1 & 0 & ... & 0 \end{pmatrix}, \; \mathbf{A}^{[\mathbf{Y},i]} = \begin{pmatrix} 0 & ... & 0 \\ & ... & \\ 0 & ... & 0 \end{pmatrix} \tag{18}$$

On a ici une contrainte d'égalité donc :

$$b_l = u_l = 0 \tag{19}$$

Le relais 4x50 4N doit être composé de deux hommes et deux femmes

$$\forall i \in [\![0; n-1]\!], \;\; 2 \cdot \left(\sum_j^m y_{i,j} \right) - p_i \leq 1$$

$$\mathbf{A}^{[\mathbf{X},i]} = \begin{pmatrix} 0 & ... & 0 \\ & ... & \\ 0 & ... & 0 \end{pmatrix}, \ \mathbf{A}^{[\mathbf{P},i]} = \begin{pmatrix} 0 & ... & 0 & -1 & 0 & ... & 0 \end{pmatrix}, \\ \mathbf{A}^{[\mathbf{Y},i]} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ & & ... & \\ 2 & 2 & 2 & 2 \\ & & ... & \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{20}$$

On a ici une contrainte d'inégalité donc :

$$b_l^{[i]} = -1, \ \ u_l^{[i]} = 1 \tag{21}$$

$$\forall j \in [\![0;3]\!], \ \ \sum_i Y_{i,j} = 1$$

$$\mathbf{A}^{[\mathbf{X},j]} = \begin{pmatrix} 0 & ... & 0 \\ & ... & \\ 0 & ... & 0 \end{pmatrix}, \ \mathbf{A}^{[\mathbf{P},j]} = \begin{pmatrix} 0 & ... & 0 \end{pmatrix}, \\ \mathbf{A}^{[\mathbf{Y},j]} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ & & ... & \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{22}$$

On a ici une contrainte d'égalité donc :

$$b_l^{[j]} = u_l^{[j]} = 1 \tag{23}$$

$$\sum_j \sum_i Y_{i,j} g_i = 0$$

$$\mathbf{A}^{[\mathbf{X},j]} = \begin{pmatrix} 0 & ... & 0 \\ & ... & \\ 0 & ... & 0 \end{pmatrix}, \ \mathbf{A}^{[\mathbf{P},j]} = \begin{pmatrix} 0 & ... & 0 \end{pmatrix}, \\ \mathbf{A}^{[\mathbf{Y},j]} = \begin{pmatrix} g_0 & \cdots & g_0 \\ & & ... \\ g_n & \cdots & g_n \end{pmatrix} \tag{24}$$

On a ici une contrainte d'égalité donc :

$$b_l = u_l = 0 \tag{25}$$

5 Document ressources

— Khandelwal, S. N. (2019). Building Teams of Experts using Integer Linear Programming. University of Windsor : lien
— MILP en python : lien
— Données sous forme de fichier .csv pour le calcul des points : lien