Optimisation convexe pour l'apprentissage de réseaux de régulation de gènes

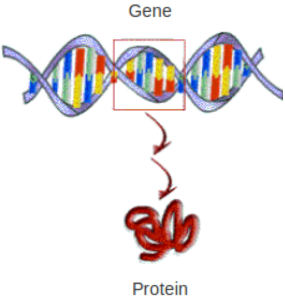
Magali Champion

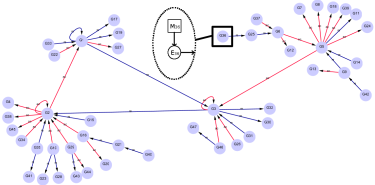
Sébastien Gadat, Christine Cierco-Ayrolles et Matthieu Vignes

12 septembre 2013

Plan

Introduction biologique...



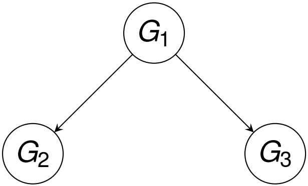


Objectif : Reconstruire le réseau de régulation de gènes $\mathcal{G} = (V, E)$ défini par :

- un noeud $i \in V$ est l'un des p gènes considérés
- une arête $(i, j) \in E$ représente une interaction entre le gène i et le gène j.

... et statistique

- p gènes étudiés, pour lesquels on connaît les données d'expression,
- échantillon de taille n.



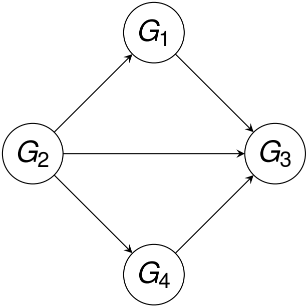
Objectif : Reconstruire le réseau de régulation de gènes $\mathcal{G} = (V, E)$ défini par :

- un noeud $i \in V$ est l'un des p gènes considérés
- une arête $(i, j) \in E$ représente une interaction entre le gène i et le gène j.

Modèle I

La première idée consiste à considérer à tour de rôle chacun des gènes G^j comme une observation dépendant de variables explicatives (les autres gènes).

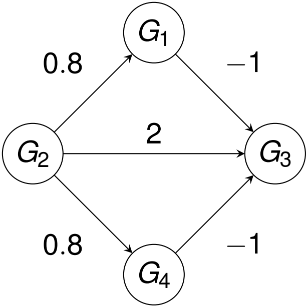
$$\forall 1 \leq j \leq p, \quad G_j = \sum_{1 \leq i \neq j \leq p} G_i + \varepsilon.$$



Modèle I

La première idée consiste à considérer à tour de rôle chacun des gènes G_j comme une observation dépendant de variables explicatives (les autres gènes).

$$\forall 1 \leq j \leq p, \quad G_j = \sum_{1 \leq i \neq j \leq p} \theta_i^j G_i + \varepsilon.$$



$\Theta = (\theta^1, ..., \theta^p)$ est la matrice d'adjacence associée au graphe $\mathcal{G}$.

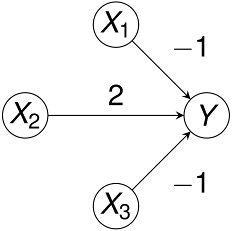
$$\Theta = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0.8 & 0 & 2 & 0.8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

Régression linéaire et boosting

On peut réécrire ce modèle sous la forme :

$$Y = X\theta + \varepsilon.$$

$\rightarrow$ Utilisation de méthodes classiques de régression du type lasso, elastic net...



Les algorithmes de type boosting permettent de construire une approximation globale de $f(X) = X\theta$ à l'aide d'approximations locales.

Algorithm 1: Algorithme $\mathbb{L}_2$-Boosting

Input: Observations $(X_i, Y_i)_{1 \leq i \leq n}$, $\gamma \in]0,1]$ et $k_{up} \in \mathbb{N}^*$

Output: $\hat{G}_{k_{up}}(f)$

Initialization : $\hat{G}_0(f) = 0$.

for $k = 1$ *to* k_{up} **do**

1. Choisir $\varphi_k \in (X_1, ..., X_{p-1})$ tel que
$$|\langle Y - \hat{G}_{k-1}(f), \varphi_k \rangle_n| = \max_{1 \leq i \leq p-1} |\langle Y - \hat{G}_{k-1}(f), X_i \rangle_n|,$$
où $\langle h_1, h_2 \rangle_n = \frac{1}{n} \sum_{i=1}^n h_1(X_i) h_2(X_i)$,
2. Calculer l'approximation courante de f
$$\hat{G}_k(f) = \hat{G}_{k-1}(f) + \gamma \langle Y - \hat{G}_{k-1}(f), \varphi_k \rangle_n \varphi_k.$$

end

Consistance de l'algorithme

1. $p_n = \underset{n \to +\infty}{O} \left(\exp(Cn^{1-\xi})\right)$, où $\xi \in (0,1)$, $C < \infty$,
2. $(\varepsilon_i)_i$ sont des v.a centrées, indépendantes de $(X_i)_i$ telles que $\mathbb{E}|\varepsilon|^t < \infty$ pour $t > \frac{4}{\xi}$,
3. $(\theta_i)_{1 \leq i \leq p_n}$ est S-parcimonieuse, de support $\mathcal{S}$, et satisfait

$$\forall j \in \mathcal{S}, \quad |\theta_j| \geq n^{-\kappa\xi}, \quad \text{où} \quad 0 < \kappa < 1.$$

Théorème (Bühlmann, 2006)

Sous les hypothèses 1 et 2, il existe $k_n := C \log n$ avec $C < \xi/4 \log(3)$ tel que :

$$\|f - \hat{G}_{k_n}(f)\| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

Recouvrement du support

On définit la cohérence ρ de la famille $(X_i)_{1 \leq i \leq p-1}$ de la manière suivante :

$$\rho = \max_{1 \leq i \neq j \leq p-1} \langle X_i, X_j \rangle.$$

Notons $\mathcal{S}_k$ le support de $\hat{G}_k(f)$.

Théorème

On suppose que $\rho(2S-1) < 1$. Sous les hypothèses 1 et 2, si γ est suffisamment petit, avec grande probabilité, on a alors :

$$\forall k, \quad \mathcal{S}_k \subset \mathcal{S}.$$

De plus, si l'hypothèse 3 est satisfaite, à la fin des itérations de l'algorithme, on a

$$\mathcal{S}_{k_n} = \mathcal{S}.$$

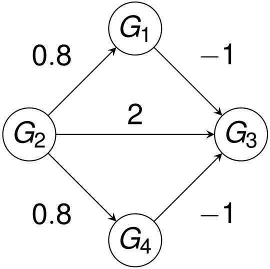
Modèle II

On considère l'ensemble des graphes acycliques orientés gaussiens (*Directed Acyclic Graphs*).

Proposition

Toute matrice d'adjacence Θ associée à un DAG $\mathcal{G}$ satisfait :

$$\Theta = PT^tP \quad \text{où} \quad \begin{cases} P & \text{matrice de permutation} \\ T & \text{matrice triangulaire inférieure stricte} \end{cases}$$



$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 2 & 0.8 & 0.8 & 0 \end{pmatrix}$$

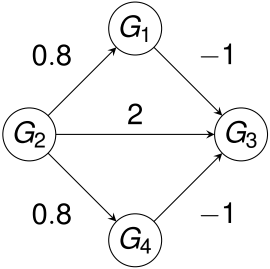
Modèle II

On considère l'ensemble des graphes acycliques orientés gaussiens (*Directed Acyclic Graphs*).

Proposition

Toute matrice d'adjacence Θ associée à un DAG $\mathcal{G}$ satisfait :

$$\Theta = PT^tP \quad \text{où} \quad \begin{cases} P & \text{matrice de permutation} \\ T & \text{matrice triangulaire inférieure stricte} \end{cases}$$



$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Problème d'optimisation

On souhaite maximiser la log-vraisemblance pénalisée :

$$(\hat{P}, \hat{T}) = \underset{P \in \mathbb{P}_p(\mathbb{R}), T \in \mathbb{T}_p(\mathbb{R})}{\operatorname{argmin}} \left\{ \frac{1}{n} \left\| G - GPT^tP \right\|_F^2 + \lambda \|T\|_1 \right\},$$

où

- $\mathbb{P}_p(\mathbb{R})$ est l'ensemble des matrices de permutation,
- $\mathbb{T}_p(\mathbb{R})$ est l'ensemble des matrices triangulaires inférieures strictes,
- $\|M\|_F^2 = \text{Trace}({}^tMM) = \sum_{i,j}(M_i^j)^2$,
- $\|M\|_1 = \sum_{i,j} |M_i^j|$.

A P fixé

Critère à minimiser :

$$\hat{T} = \underset{T \in \mathbb{T}_p(\mathbb{R})}{\operatorname{argmin}} \left\{ \frac{1}{n} \left\| G - GPT^t P \right\|_F^2 + \lambda \| T \|_1 \right\}.$$

- minimisation d'une fonction convexe, différentiable et quadratique + un terme de pénalisation

$$T_{k+1} = \underset{T}{argmin} \left\{ \frac{L}{2} \left\| T - \left(T_k - \frac{\nabla f(T_k)}{L} \right) \right\|_F^2 + \lambda \| T \|_1 \right\}.$$

- projection sur l'espace des contraintes $\mathbb{T}_p(\mathbb{R})$.

A T fixé

Critère à minimiser : $\hat{P} = \underset{P \in \mathbb{P}_p(\mathbb{R})}{\operatorname{argmin}} \left\{ \frac{1}{n} \left\| G - GPT^t P \right\|_F^2 \right\}.$

Comme l'espace des contraintes n'est pas convexe, on propose une relaxation convexe du critère à minimiser.

Definition

Une matrice bistochastique $A = (a_{ij})_{1 \leq i,j \leq p}$ est une matrice telle que :

- $a_{ij} \geq 0$,
- $\sum_i a_{ij} = \sum_j a_{ij} = 1$.

Proposition (Birkhoff)

L'ensemble des matrices bistochastiques $\mathbb{B}_p(\mathbb{R})$ est un ensemble convexe dont les matrices de permutation sont les points extrémaux.

Minimisation alternée

$$(\hat{P}, \hat{T}) = \underset{P \in \mathbb{B}_p(\mathbb{R}), T \in \mathbb{T}_p(\mathbb{R})}{\operatorname{argmin}} \frac{1}{n} \left\| G - GPT^t P \right\|_F^2 + \lambda \|T\|_1 .$$

$$P_0 \in \mathbb{P}_p(\mathbb{R}) \xrightarrow{\text{optimisation}} T_0 \xrightarrow{\text{proj}} T_0' \in \mathbb{T}_p(\mathbb{R})$$

descente de gradient projeté

$$P_1 \in \mathbb{B}_p(\mathbb{R}) \xrightarrow{\text{optimisation}} T_1 \xrightarrow{\text{proj}} T_1' \in \mathbb{T}_p(\mathbb{R})$$

descente de gradient projeté

$$P_2 \in \mathbb{B}_p(\mathbb{R}) \longrightarrow \cdots \xrightarrow{\text{Projection sur } \mathbb{P}_p(\mathbb{R})}$$

Matrices de permutation

On veut trouver la projection d'une matrice bistochastique $B \in \mathbb{B}_p(\mathbb{R})$ sur $\mathbb{P}_p(\mathbb{R})$, ce qui s'écrit sous la forme :

$$\begin{aligned} \mathrm{Proj}_{\mathbb{P}_p(\mathbb{R})}(B) &= \underset{P \in \mathbb{P}_p(\mathbb{R})}{\mathrm{argmin}} \ \|B - P\|_F \\ &= \underset{P \in \mathbb{P}_p(\mathbb{R})}{\mathrm{argmin}} - 2\langle B, P \rangle_F. \end{aligned}$$

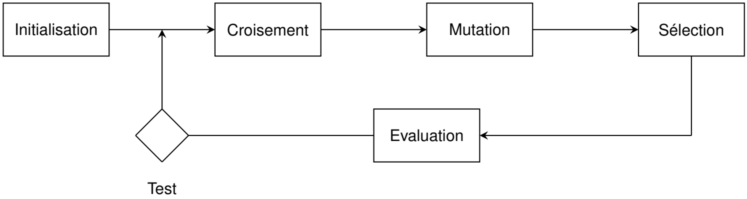
La fonction $-2\langle B, P \rangle_F$ à minimiser est linéaire et l'espace des contraintes $\mathbb{P}_p(\mathbb{R})$ est l'ensemble des points extrémaux d'un polyèdre convexe. Il existe donc un point extrémal solution du problème relaxé

$$P = \underset{P \in \mathbb{B}_p(\mathbb{R})}{\mathrm{argmin}} - 2\langle B, P \rangle_F.$$

Algorithmes génétiques

Travail en collaboration avec Victor Pichény

Au lieu de relaxer la condition $P \in \mathbb{P}_p(\mathbb{R})$, nous proposons d'utiliser des algorithmes génétiques, qui sont des procédés de recherche heuristiques reproduisant les processus d'évolution naturelle.



En quelques mots (initialisation)

1. On considère N matrices de permutation. Chacune d'entre elles est représentée par une suite de "gènes", appelée "chromosome".

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

chromosome $\longrightarrow$

2	3	6	1	4	5

2. On cherche la matrice triangulaire inférieure stricte T associée à chaque chromosome.

En quelques mots (croisement)

1. Sélection pour le croisement : roulette wheel selection
2. Méthode de croisement

1	2	3	4	5	6
6	5	1	3	2	4

$\longrightarrow$

		1	3	2	
		3	4	5	

En quelques mots (croisement)

1. Sélection pour le croisement : roulette wheel selection
2. Méthode de croisement

1	2	3	4	5	6
6	5	1	3	2	4

$\longrightarrow$

4	5	1	3	2	6
6	1	3	4	5	2

Résultats numériques

Pour comparer les calculs, on calcule :

- la MSE : $\left\|\hat{\Theta} - \Theta^*\right\|_F^2$
- la MSEP : $\frac{1}{n}\left\|G - G\hat{\Theta}\right\|_F^2$.

Pour $n = 100$ et $p = 5$

	Optimisation	Algorithmes génétiques
Temps de calcul	10 s	11 s
MSE	2.45	2.10
MSEP	3.74	3.72

Résultats numériques

Pour comparer les calculs, on calcule :

- la MSE : $\left\|\hat{\Theta} - \Theta^*\right\|_F^2$
- la MSEP : $\frac{1}{n}\left\|G - G\hat{\Theta}\right\|_F^2$.

Pour $n = 100$ et $p = 50$

	Optimisation	Algorithmes génétiques
Temps de calcul	26 s	15 s
MSE	13.20	6.18
MSEP	20.82	12.44