Contribution à la modélisation et l'inférence de réseaux de régulation de gènes

Magali Champion

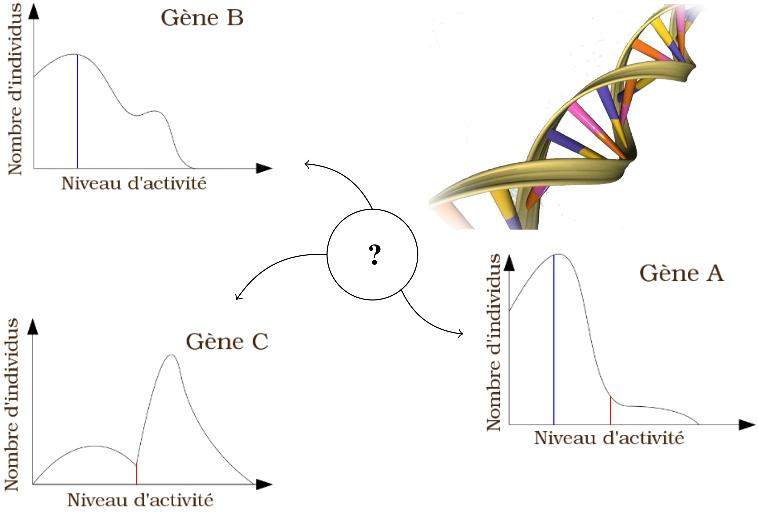
Sous la direction de Sébastien Gadat (IMT),
Christine Cierco-Ayrolles et Matthieu Vignes (INRA)

INSTITUT de MATHEMATIQUES de TOULOUSE

INRA SCIENCE & IMPACT

05 décembre 2014

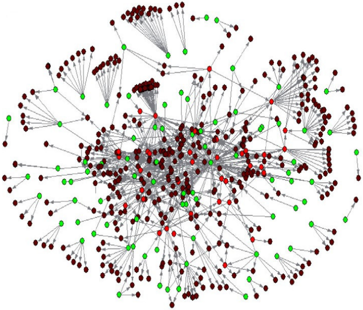
Motivation biologique



Objectif :

Retrouver le graphe $\mathcal{G}$ qui modélise les interactions entre les p gènes étudiés à partir de deux types de données prises sur un échantillon de taille n :

- des données continues $E^1, ..., E^p$,
- des données discrètes $M^1, ..., M^p$.

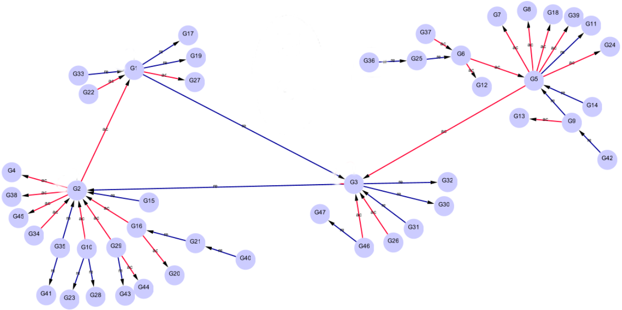


On appelle réseau de régulation de gènes un tel graphe.

Cadre d'étude

Un réseau de régulation de gènes peut être modélisé par (Liu *et al.*, 2008) :

$$\forall j \in [\![1, p]\!], \quad E^j = \sum_{i=1}^{p} B_i^j E^i + \sum_{i=1}^{p} A_i^j M^i + \varepsilon^j.$$



Cadre d'étude

Un réseau de régulation de gènes peut être modélisé par (Liu *et al.*, 2008) :

$$\forall j \in [\![1, p]\!], \;\; E^j = \sum_{i=1}^{p} B_i^j E^i + \sum_{i=1}^{p} A_i^j M^i + \varepsilon^j.$$

Une méthode classique pour résoudre ce type de problèmes consiste à effectuer p régressions indépendantes à l'aide de méthodes pénalisées du type :

- Lasso (Tibshirani, 1996), ajout d'une pénalité en norme ℓ_1,
- Dantzig Selector (Candes *et al.*, 2007),

Cadre d'étude

Un réseau de régulation de gènes peut être modélisé par (Liu *et al.*, 2008) :

$$\forall j \in [\![1,p]\!], \quad E^j = \sum_{i=1}^{p} B_i^j E^i + \sum_{i=1}^{p} A_i^j M^i + \varepsilon^j.$$

Une méthode classique pour résoudre ce type de problèmes consiste à effectuer p régressions indépendantes à l'aide de méthodes pénalisées du type :

- Lasso (Tibshirani, 1996), ajout d'une pénalité en norme ℓ_1,
- Dantzig Selector (Candes *et al.*, 2007),
- Autres formalismes :

  Forêts aléatoires (Huynh *et al.*, 2010),
  Réseaux bayésiens (Friedman *et al.*, 2000).

Contributions

1. Algorithmes de $\mathbb{L}_2$-Boosting et leurs applications
   - Les algorithmes de $\mathbb{L}_2$-Boosting pour la résolution de régressions parcimonieuses multi-tâches
   - Application des algorithmes de $\mathbb{L}_2$-Boosting à l'analyse de sensibilité pour l'estimation d'indices de Sobol

2. Apprentissage de liens causaux dans des graphes acycliques dirigés
   - Estimation de graphe par maximum de vraisemblance pénalisé
   - Inégalités en prédiction et en estimation
   - Implémentation d'un algorithme génétique hybride

1. Algorithmes de $\mathbb{L}_2$-Boosting et leurs applications
   - Les algorithmes de $\mathbb{L}_2$-Boosting pour la résolution de régressions parcimonieuses multi-tâches
   - Application des algorithmes de $\mathbb{L}_2$-Boosting à l'analyse de sensibilité pour l'estimation d'indices de Sobol

2. Apprentissage de liens causaux dans des graphes acycliques dirigés
   - Estimation de graphe par maximum de vraisemblance pénalisé
   - Inégalités en prédiction et en estimation
   - Implémentation d'un algorithme génétique hybride

Régressions parcimonieuses

On considère un échantillon d'apprentissage $(X_k, Y_k)_{1 \leq k \leq n}$ constitué de n copies i.i.d. de $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^m$ avec :

$$Y = f(X) + \varepsilon,$$

où ε est une variable aléatoire centrée indépendante de X modélisant la présence de bruit, et $f := (f^1, \ldots, f^m)$ est inconnue.

Régressions parcimonieuses

On considère un échantillon d'apprentissage $(X_k, Y_k)_{1\leq k\leq n}$ constitué de n copies i.i.d. de $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^m$ avec :

$$Y = f(X) + \varepsilon,$$

où ε est une variable aléatoire centrée indépendante de X modélisant la présence de bruit, et $f := (f^1, ..., f^m)$ est inconnue.

Objectif : estimer f à partir des données.

Régressions parcimonieuses

On considère un échantillon d'apprentissage $(X_k, Y_k)_{1\leq k\leq n}$ constitué de n copies i.i.d. de $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^m$ avec :

$$Y = f(X) + \varepsilon,$$

où ε est une variable aléatoire centrée indépendante de X modélisant la présence de bruit, et $f := (f^1, ..., f^m)$ est inconnue.

Hypothèse : il existe un système de fonctions $(g_j)_{1\leq j\leq p}$ normées, noté $\mathcal{D}$ et $(a_{i,j})_j$, S^i-parcimonieuse de support $\mathcal{S}^i$, tels que :

$$\forall i \in [\![1, m]\!],\ \ f^i(X) = \sum_{j=1}^{p} a_{i,j} g_j(X^i).$$

Objectif : estimer f à partir des données.

Régressions parcimonieuses

On considère un échantillon d'apprentissage $(X_k, Y_k)_{1 \leq k \leq n}$ constitué de n copies i.i.d. de $(X, Y) \in \mathbb{R}^p \times \mathbb{R}^m$ avec :

$$Y = f(X) + \varepsilon,$$

où ε est une variable aléatoire centrée indépendante de X modélisant la présence de bruit, et $f := (f^1, \ldots, f^m)$ est inconnue.

Hypothèse : il existe un système de fonctions $(g_j)_{1 \leq j \leq p}$ normées, noté $\mathcal{D}$ et $(a_{i,j})_j$, S^i-parcimonieuse de support $\mathcal{S}^i$, tels que :

$$\forall i \in [\![1, m]\!], \;\; f^i(X) = \sum_{j=1}^{p} a_{i,j} g_j(X^i).$$

Objectif : estimer f à partir des données.

$$\forall i \in [\![1, m]\!], \;\; \hat{f}^i(X) = \sum_{j=1}^{p} \hat{a}_{i,j} g_j(X^i).$$

Les algorithmes de $\mathbb{L}_2$-Boosting ont pour but d'approcher la fonction f par une succession d'approximations locales. Notons :

- $\hat{G}_k(f)$ l'approximation de la fonction f à l'étape k,
- $\hat{R}_k(f)$ le résidu courant observable : $\hat{R}_k(f) := Y - \hat{G}_k(f)$.

Algorithme $\mathbb{L}_2$-Boosting

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

On choisit $\varphi_k \in \mathcal{D}$ tel que

$$\left|\langle \hat{R}_{k-1}(f), \varphi_k \rangle_n\right| = \max_{1 \leq j \leq p_n} \left|\langle \hat{R}_{k-1}(f), g_j \rangle_n\right|,$$

On calcule la nouvelle approximation et le résidu associés à f

$$\hat{G}_k(f) = \hat{G}_{k-1}(f) + \gamma \langle \hat{R}_{k-1}(f), \varphi_k \rangle_n \varphi_k.$$

End

Extension au cas multi-tâches

Pour étendre l'algorithme, le point crucial est de choisir à la fois la coordonnée du résidu qui nécessite le plus d'amélioration, et l'élément du dictionnaire le plus pertinent $\varphi_k \in \mathcal{D}$.

- Sélection du prédicteur :

$$\sum_{j=1}^{p} \langle \hat{R}_{k-1}(f^{i_k}), g_j \rangle_n^2 = \max_{1 \leq i \leq m} \sum_{j=1}^{p} \langle \hat{R}_{k-1}(f^{i}), g_j \rangle_n^2,$$

- Sélection de l'élément du dictionnaire :

$$\left| \langle \hat{R}_{k-1}(f^{i_k}), \varphi_k \rangle_n \right| = \max_{1 \leq j \leq p} \left| \langle \hat{R}_{k-1}(f^{i_k}), g_j \rangle_n \right|,$$

- Calcul de l'approximation courante :

$$\hat{G}_k(f^{i_k}) = \hat{G}_{k-1}(f^{i_k}) + \gamma \langle \hat{R}_{k-1}(f^{i_k}), \varphi_k \rangle_n \varphi_k.$$

Résultats de consistance ($m = 1$)

- **$\mathbf{H}_{\mathbf{dim}}$** : $$p_n = \underset{n \to +\infty}{\mathcal{O}} \left(\exp\left(n^{1-\xi}\right)\right), \text{ où } \xi \in (0,1),$$
- **$\mathbf{H}_{\varepsilon}$** : $$\sup_{1 \le i \le m_n} \mathbb{E}\left|\varepsilon^i\right|^t < \infty, \text{ pour un } t > \frac{4}{\xi}.$$

Théorème (Consistance, Bühlmann (2006))

Sous les hypothèses $\mathbf{H}_{\mathbf{dim}}$ et $\mathbf{H}_{\varepsilon}$, il existe $k_n := C\log(n)$ telle que si $S\sqrt{\frac{\log p}{n}} \xrightarrow[n \to +\infty]{} 0$:

$$\mathbb{E}\left\|f - \hat{G}_{k_n}(f)\right\|_n^2 \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

Résultats de consistance ($m = 1$)

- $\mathbf{H_{dim}}$: $$p_n = \underset{n\to+\infty}{\mathcal{O}}\left(\exp\left(n^{1-\xi}\right)\right), \text{ où } \xi \in (0,1),$$
- $\mathbf{H}_\varepsilon$: $$\sup_{1\leq i\leq m_n} \mathbb{E}\left|\varepsilon^i\right|^t < \infty, \text{ pour un } t > \frac{4}{\xi}.$$

Théorème (Consistance, Bühlmann (2006))

Sous les hypothèses $\mathbf{H_{dim}}$ *et* $\mathbf{H}_\varepsilon$*, il existe* $k_n := C\log(n)$ *telle que si* $S\sqrt{\frac{\log p}{n}} \xrightarrow[n\to+\infty]{} 0$ *:*

$$\mathbb{E}\left\|f - \hat{G}_{k_n}(f)\right\|_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Résultats de consistance

- $\mathbf{H}^{\mathbf{Mult}}_{\mathbf{dim}}$: $m_n = \underset{n\to+\infty}{\mathcal{O}}\left(\exp\left(n^{1-\xi}\right)\right)$, $p_n = \underset{n\to+\infty}{\mathcal{O}}\left(n^{\xi/4}\right)$ où $\xi \in (0,1)$,
- $\mathbf{H}^{\mathbf{Mult}}_{\varepsilon}$: $\sup_{1\le i\le m_n} \mathbb{E}\left|\varepsilon^i\right|^t < \infty$, $t > \frac{4}{\xi}$, et $\forall i \neq j$, $\mathrm{Var}\left(\varepsilon^j\right) = \mathrm{Var}\left(\varepsilon^i\right)$,
- $\mathbf{H}_{\mathbf{RE}^{\pm}}$: La plus petite et la plus grande valeur propre λ_{min} et λ_{max} de ${}^tD_{\mathcal{S}^i}D_{\mathcal{S}^i}$ sont indépendantes de n telles que :

$$0 < \lambda_{min} < \ldots < \lambda_{max} < \infty.$$

Théorème (Consistance, Champion *et al.* (2014))

Sous les hypothèses $\mathbf{H}^{\mathbf{Mult}}_{\mathbf{dim}}$, $\mathbf{H}^{\mathbf{Mult}}_{\varepsilon}$ *et* $\mathbf{H}_{\mathbf{RE}^{\pm}}$, *il existe* $k_n := C\log(n)$ *telle que si* $S^i\sqrt{\frac{\log p}{n}} \underset{n\to+\infty}{\longrightarrow} 0$:

$$\sup_{1\le i\le m_n} \mathbb{E}\left\|f^i - \hat{G}_{k_n}(f^i)\right\|_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Résultats de consistance

- $\mathbf{H_{dim}^{Mult}}$: $m_n = \underset{n\to+\infty}{\mathcal{O}}\left(\exp\left(n^{1-\xi}\right)\right)$, $p_n = \underset{n\to+\infty}{\mathcal{O}}\left(n^{\xi/4}\right)$ où $\xi \in (0,1)$,
- $\mathbf{H_{\varepsilon}^{Mult}}$: $\sup\limits_{1\leq i\leq m_n} \mathbb{E}\left|\varepsilon^i\right|^t < \infty$, $t > \frac{4}{\xi}$, et $\forall i \neq j$, $\mathrm{Var}\left(\varepsilon^j\right) = \mathrm{Var}\left(\varepsilon^i\right)$,
- $\mathbf{H_{RE^\pm}}$: La plus petite et la plus grande valeur propre λ_{min} et λ_{max} de ${}^tD_{\mathcal{S}^i}D_{\mathcal{S}^i}$ sont indépendantes de n telles que :

$$0 < \lambda_{min} < \ldots < \lambda_{max} < \infty.$$

Théorème (Consistance, Champion *et al.* (2014))

Sous les hypothèses $\mathbf{H_{dim}^{Mult}}$, $\mathbf{H_{\varepsilon}^{Mult}}$ *et* $\mathbf{H_{RE^\pm}}$, *il existe* $k_n := C\log(n)$ *telle que si* $S^i\sqrt{\frac{\log p}{n}} \xrightarrow[n\to+\infty]{} 0$:

$$\sup_{1\leq i\leq m_n} \mathbb{E}\left\|f^i - \hat{G}_{k_n}(f^i)\right\|_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Recouvrement du support

Notons $\mathcal{S}_k^i$ le support de $\hat{G}_k(f^i)$ à l'étape k.

- **$\mathbf{H_S}$** : $\max_{j \notin \mathcal{S}^i} \left\| D^+_{\mathcal{S}^i} g_j \right\|_1 < 1$, où D^+ est le pseudo-inverse de D,
- **$\mathbf{H_{SNR}}$** : $\forall j \in \mathcal{S}^i, |a_{ij}| \geq \log(n)^{-\kappa}$, avec $\kappa > 0$.

Théorème (Recouvrement du support, Champion *et al.* (2014))

Sous les hypothèses $\mathbf{H_{dim}}$, $\mathbf{H}_\varepsilon$, $\mathbf{H_{RE^{\pm}}}$ et $\mathbf{H_S}$, avec grande probabilité :

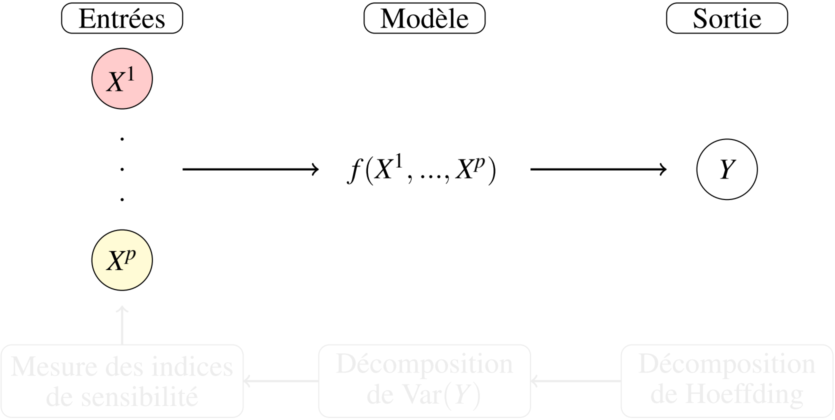
$$\forall k, \quad \mathcal{S}_k^i \subset \mathcal{S}^i.$$

De plus, si $\mathbf{H_{SNR}}$ est satisfaite pour κ suffisamment petit, avec grande probabilité, à la fin des itérations du Boosting :

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

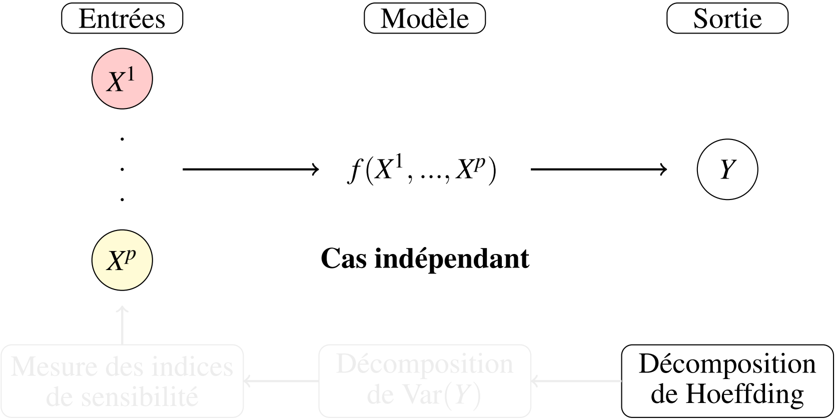
1. Algorithmes de $\mathbb{L}_2$-Boosting et leurs applications
   - Les algorithmes de $\mathbb{L}_2$-Boosting pour la résolution de régressions parcimonieuses multi-tâches
   - Application des algorithmes de $\mathbb{L}_2$-Boosting à l'analyse de sensibilité pour l'estimation d'indices de Sobol
2. Apprentissage de liens causaux dans des graphes acycliques dirigés
   - Estimation de graphe par maximum de vraisemblance pénalisé
   - Inégalités en prédiction et en estimation
   - Implémentation d'un algorithme génétique hybride

Analyse de sensibilité



Objectif : Mesurer la variabilité de la réponse en fonction des paramètres d'entrée.

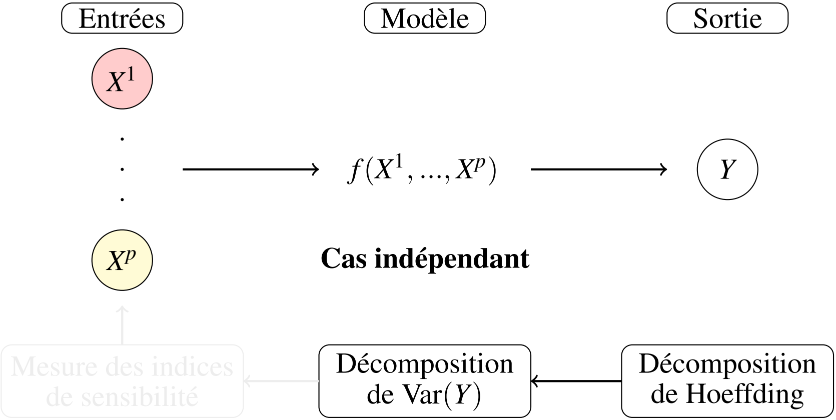
Analyse de sensibilité



$$f(X) = f_\emptyset + \sum_i f_i(X^i) + \sum_{1 \leq i < j \leq p} f_{ij}(X^i, X^j) + \ldots + f_{1,\ldots,p}(X) = \sum_{u \in S} f_u(X^u),$$

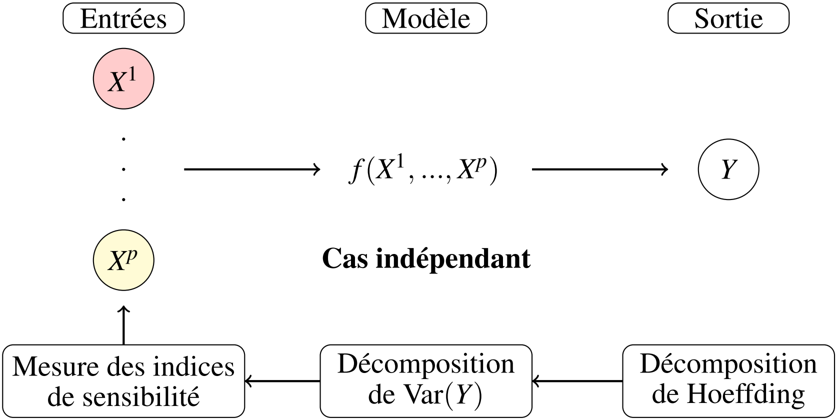
où $(f_u)_{u \in S}$ est une famille orthogonale.

Analyse de sensibilité



Alors : $\mathrm{Var}(Y) = \sum_{u \in S} \mathrm{Var}\left(f_u(X^u)\right)$.

Analyse de sensibilité



où l'indice de sensibilité d'ordre $|u|$ est : $S_u = \frac{\mathrm{Var}(f_u(X^u))}{\mathrm{Var}(Y)}$.

Décomposition de Hoeffding - cas dépendant

Chastaing *et. al* (2012)

f se décompose de manière unique sous la forme d'une somme de fonctions orthogonalement hiérarchiques :

$$f(X) = \underbrace{f_\emptyset}_{\text{constant}} + \sum_i \underbrace{f_i(X^i)}_{\text{effets principaux}} + \sum_{1 \leq i < j \leq p} \underbrace{f_{ij}(X^i, X^j)}_{\text{effets d'interaction}} + \ldots + \underbrace{f_{1,\ldots,p}(X)}_{\text{résidu}},$$

où $(f_u)_u \in H_u$ est hiérarchiquement orthogonale.

Exemple de décomposition orthogonale hiérarchique

$f_{ij}(X^i, X^j) \in H_{ij}$ vérifie $\left\langle f_{ij}(X^i, X^j), f_j(X^j) \right\rangle = \left\langle f_{ij}(X^i, X^j), f_i(X^i) \right\rangle = 0.$

Procédure d'orthogonalisation hiérarchique de Gram-Schmidt

Cette procédure consiste à construire une base de l'espace H_u. On considère uniquement les effets d'interactions d'ordre $|u| \leq 2$.

1. Initialisation : pour tout i, on considère une base $(\phi_k^i)_{k=0}^L$ de $L^2(X^i)$. Alors :

$$H_i^L = \text{Vect}\{1, \phi_1^i, ..., \phi_L^i\}.$$

2. Pour tout (i, j), étant donnés H_i^L et H_j^L, on construit H_{ij}^L :

$$\begin{array}{ll} & \text{une combinaison linéaire des éléments de } H_i^L \rightarrow \lambda_1 \\ + & \text{une combinaison linéaire des éléments de } H_j^L \rightarrow \lambda_2 \\ + & \phi^i(X^i) \times \phi^j(X^j) \end{array}$$

On vérifie que (λ_1, λ_2) satisfait les contraintes :

$$A^{ij} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = D^{ij}.$$

Procédure d'orthogonalisation hiérarchique de Gram-Schmidt

Cette procédure consiste à construire une base de l'espace H_u. On considère uniquement les effets d'interactions d'ordre $|u| \leq 2$.

1. Initialisation : pour tout i, on considère une base $(\phi_k^i)_{k=0}^L$ de $L^2(X^i)$. Alors :

$$H_i^L = \text{Vect}\{1, \phi_1^i, ..., \phi_L^i\}.$$

2. Pour tout (i,j), étant donnés H_i^L et H_j^L, on construit H_{ij}^L :

$$\begin{array}{ll} & \text{une combinaison linéaire des éléments de } H_i^L \rightarrow \lambda_1 \\ + & \text{une combinaison linéaire des éléments de } H_j^L \rightarrow \lambda_2 \\ + & \phi^i(X^i) \times \phi^j(X^j) \end{array}$$

On vérifie que (λ_1, λ_2) satisfait les contraintes :

$$A^{ij} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = D^{ij}.$$

Comment estimer cette décomposition ?

$$\begin{aligned} f(X) &= \underbrace{f_\emptyset}_{\text{terme constant}} + \sum_i \underbrace{f_i(X^i)}_{\text{effets principaux}} \\ &+ \sum_{1\leq i<j\leq p} \underbrace{f_{ij}(X^i,X^j)}_{\text{effets d'interaction}} + \ldots + \underbrace{f_{1,\ldots,p}(X)}_{\text{résidu}}, \end{aligned}$$

Deux problématiques distinctes :

1. approximation des bases de H_u,
2. estimation parcimonieuse des coefficients dans cette base afin de sélectionner les interactions principales.

Greedy Hierarchically Orthogonal Functional Decomposition

Initialisation : On coupe $\mathcal{O}$ en deux $\mathcal{O}_1 \cup \mathcal{O}_2$ de taille (n_1, n_2).

Step 1 : $\forall u \in S$, on utilise la procédure d'orthogonalisation hiérarchique de Gram-Schmidt empirique avec $\mathcal{O}_1$ pour constuire une approximation de $H_u^{L,0}$:

$$\hat{H}_u^{L,0,n_1} := \text{Vect}\{\hat{\phi}_{1,n_1}^u, \cdots, \hat{\phi}_{L,n_1}^u\}.$$

Step 2 : On utilise du $\mathbb{L}_2$-Boosting avec $\mathcal{O}_2$ sur le dictionnaire appris $\mathcal{D} = \{\hat{\phi}_{1,n_1}^1, \cdots \hat{\phi}_{L,n_1}^1, \cdots, \hat{\phi}_{1,n_1}^u, \cdots, \hat{\phi}_{L,n_1}^u, \cdots\}$ pour obtenir une décomposition parcimonieuse.

Greedy Hierarchically Orthogonal Functional Decomposition

Initialisation : On coupe $\mathcal{O}$ en deux $\mathcal{O}_1 \cup \mathcal{O}_2$ de taille (n_1, n_2).

Step 1 : $\forall u \in S$, on utilise la procédure d'orthogonalisation hiérarchique de Gram-Schmidt empirique avec $\mathcal{O}_1$ pour constuire une approximation de $H_u^{L,0}$:

$$\hat{H}_u^{L,0,n_1} := \text{Vect}\{\hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L,n_1}\}.$$

Step 2 : On utilise du $\mathbb{L}_2$-Boosting avec $\mathcal{O}_2$ sur le dictionnaire appris $\mathcal{D} = \{\hat{\phi}^1_{1,n_1}, \cdots \hat{\phi}^1_{L,n_1}, \cdots, \hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L,n_1}, \cdots\}$ pour obtenir une décomposition parcimonieuse.

$$f(X) = \sum_{\substack{u \in S^* \\ |u| \leq d}} \sum_{l_u} \beta^u_{l_u} \phi^u_{l_u}(X^u) \simeq \sum_{\substack{u \in S^* \\ |u| \leq d}} \sum_{l_u} \beta^u_{l_u} \hat{\phi}^u_{l_u,n_1}(X^u).$$

Greedy Hierarchically Orthogonal Functional Decomposition

Initialisation : On coupe $\mathcal{O}$ en deux $\mathcal{O}_1 \cup \mathcal{O}_2$ de taille (n_1, n_2).

Step 1 : $\forall u \in S$, on utilise la procédure d'orthogonalisation hiérarchique de Gram-Schmidt empirique avec $\mathcal{O}_1$ pour constuire une approximation de $H_u^{L,0}$:

$$\hat{H}_u^{L,0,n_1} := \text{Vect}\{\hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L,n_1}\}.$$

Step 2 : On utilise du $\mathbb{L}_2$-Boosting avec $\mathcal{O}_2$ sur le dictionnaire appris $\mathcal{D} = \{\hat{\phi}^1_{1,n_1}, \cdots \hat{\phi}^1_{L,n_1}, \cdots, \hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L,n_1}, \cdots\}$ pour obtenir une décomposition parcimonieuse.

$$f(X) = \sum_{\substack{u \in S^* \\ |u| \leq d}} \sum_{l_u} \beta^u_{l_u} \phi^u_{l_u}(X^u) \simeq \sum_{\substack{u \in S^* \\ |u| \leq d}} \sum_{l_u} \beta^u_{l_u} \hat{\phi}^u_{l_u,n_1}(X^u).$$

Greedy Hierarchically Orthogonal Functional Decomposition

Initialisation : On coupe $\mathcal{O}$ en deux $\mathcal{O}_1 \cup \mathcal{O}_2$ de taille (n_1, n_2).

Step 1 : $\forall u \in S$, on utilise la procédure d'orthogonalisation hiérarchique de Gram-Schmidt empirique avec $\mathcal{O}_1$ pour constuire une approximation de $H_u^{L,0}$:

$$\hat{H}_u^{L,0,n_1} := \text{Vect}\{\hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L,n_1}\}.$$

Step 2 : On utilise du $\mathbb{L}_2$-Boosting avec $\mathcal{O}_2$ sur le dictionnaire appris $\mathcal{D} = \{\hat{\phi}^1_{1,n_1}, \cdots \hat{\phi}^1_{L,n_1}, \cdots, \hat{\phi}^u_{1,n_1}, \cdots, \hat{\phi}^u_{L,n_1}, \cdots\}$ pour obtenir une décomposition parcimonieuse.

$$f(X) = \sum_{\substack{u \in S^* \\ |u| \le d}} \sum_{l_u} \beta^u_{l_u} \phi^u_{l_u}(X^u) \simeq \sum_{\substack{u \in S^* \\ |u| \le d}} \sum_{l_u} \hat{\beta}^u_{l_u} \hat{\phi}^u_{l_u,n_1}(X^u).$$

Convergence de la base approchée

- $\mathbf{H_b^1}$: $\sup_{i \in [\![1,p]\!]} \left\| \phi^i(X_i) \right\|_\infty < +\infty$,
- $\mathbf{H_b^2}$: $p_n = \underset{n \to +\infty}{\mathcal{O}} \left(\exp(n^{1-\xi}) \right)$, avec $0 < \xi \leq 1$,
- $\mathbf{H_b^3}$: La matrice de Gram A du système linéaire satisfait :

$$\exists C > 0, \quad \forall (i,j) \in [\![1, p_n]\!]^2, \qquad det(A^{ij}) \geq C n^{-\vartheta}.$$

Théorème (Champion *et al.* (2014))

Sous les hypothèses $\mathbf{H_b^{1-3}}$*, si* $\vartheta < \xi/2$*, la suite* $\left(\hat{\phi}^u \right)_u$ *vérifie :*

$$\sup_u \left\| \hat{\phi}^u - \phi^u \right\| = \mathcal{O}_P(n^{\vartheta - \xi/2}).$$

Consistance de l'algorithme

- $\mathbf{H}_\varepsilon$: $\mathbb{E}\,|\varepsilon|^t < \infty$, pour un certain $t > 4/\xi$,

Notons
- $S := \sum_{\substack{u \in S^* \\ |u| \leq d}} \sum_{l_u} \mathbb{1}_{\beta^u_{l_u} \neq 0}$ la parcimonie de f,
- $\hat{f}(X) = \sum_{\substack{u \in S^* \\ |u| \leq d}} \sum_{l_u} \hat{\beta}^u_{l_u} \hat{\phi}^u_{l_u, n_1}$.

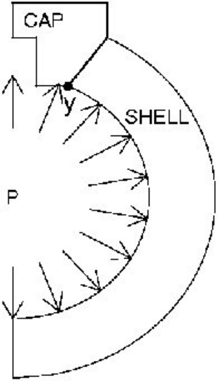
Théorème (Consistance, Champion *et al.* (2014))

Sous les hypothèses $\mathbf{H}_\mathbf{b}^{\mathbf{1-3}}$ *et* $\mathbf{H}_\varepsilon$, *si* $\vartheta < \xi/2$, *il existe* $k_n := C \log n$ *telle que, si* $S\sqrt{\frac{\log p}{n}} \xrightarrow[n \to \infty]{} 0$ *:*

$$\left\| \hat{f} - f \right\| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

Applications numériques

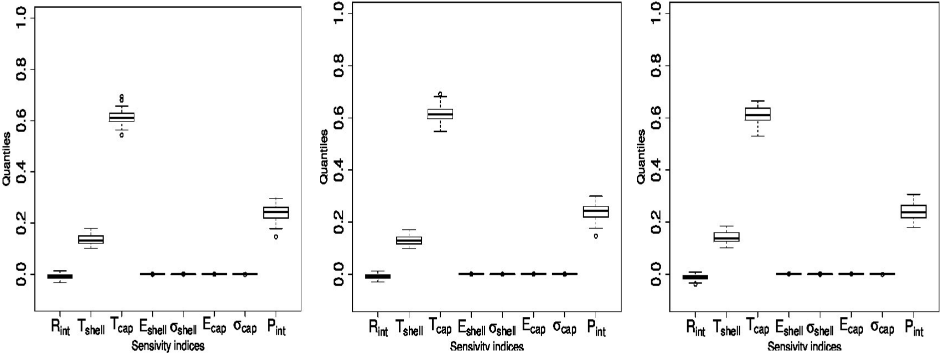
Variables dépendantes



Le critère de von Mises (ou critère énergétique de distorsion élastique) est à expliquer en fonction des variables suivantes :

- le rayon interne de la paroi R_{int},
- l'épaisseur de la paroi T_{shell},
- l'épaisseur du bouchon T_{cap},
- la pression interne P_{int},
- le module de Young E_{shell}, E_{cap},
- la limite d'élasticité σ_{shell} et σ_{cap}.

Résultats numériques



(a) FoBa (b) Boosting (c) Lasso

Méthodes	$\|\hat{\beta}\|_0$	Temps de calcul (en s)
$\mathbb{L}_2$-Boosting	10	0.0266
FoBa	22	0.3741
Lasso	23	0.15

1. Algorithmes de $\mathbb{L}_2$-Boosting et leurs applications
   - Les algorithmes de $\mathbb{L}_2$-Boosting pour la résolution de régressions parcimonieuses multi-tâches
   - Application des algorithmes de $\mathbb{L}_2$-Boosting à l'analyse de sensibilité pour l'estimation d'indices de Sobol

2. Apprentissage de liens causaux dans des graphes acycliques dirigés
   - Estimation de graphe par maximum de vraisemblance pénalisé
   - Inégalités en prédiction et en estimation
   - Implémentation d'un algorithme génétique hybride

1. Algorithmes de $\mathbb{L}_2$-Boosting et leurs applications
   - Les algorithmes de $\mathbb{L}_2$-Boosting pour la résolution de régressions parcimonieuses multi-tâches
   - Application des algorithmes de $\mathbb{L}_2$-Boosting à l'analyse de sensibilité pour l'estimation d'indices de Sobol

2. Apprentissage de liens causaux dans des graphes acycliques dirigés
   - Estimation de graphe par maximum de vraisemblance pénalisé
   - Inégalités en prédiction et en estimation
   - Implémentation d'un algorithme génétique hybride

Modèle et hypothèses

On s'intéresse au modèle suivant :

$$\forall j \in [\![1, p]\!], \;\; X^j = \sum_{i=1, i \neq j}^{p} (G_0)_i^j \; X^i + \varepsilon^j,$$

où $\varepsilon^j \sim \mathcal{N}(0, \sigma_j^2)$ et $(G_0)_i^j \neq 0$ représente l'effet du nœud (gène) X^j vers le nœud X^i, sous les hypothèses suivantes :

1. le graphe $\mathcal{G}_0$ associé à la matrice G_0 est un *Graphe Acyclique Dirigé* (DAG),

Modèle et hypothèses

On s'intéresse au modèle suivant :

$$\forall j \in [\![1, p]\!], \quad X^j = \sum_{i=1, i \neq j}^{p} (G_0)_i^j \; X^i + \varepsilon^j,$$

où $\varepsilon^j \sim \mathcal{N}(0, \sigma_j^2)$ et $(G_0)_i^j \neq 0$ représente l'effet du nœud (gène) X^j vers le nœud X^i, sous les hypothèses suivantes :

1. le graphe $\mathcal{G}_0$ associé à la matrice G_0 est un *Graphe Acyclique Dirigé* (DAG),
2. les variables aléatoires $(X^i)_{1 \leq i \leq p}$ suivent une loi normale,

Modèle et hypothèses

On s'intéresse au modèle suivant :

$$\forall j \in [\![1, p]\!], \ \ X^j = \sum_{i=1, i\neq j}^{p} (G_0)_i^j \ X^i + \varepsilon^j,$$

où $\varepsilon^j \sim \mathcal{N}(0, \sigma_j^2)$ et $(G_0)_i^j \neq 0$ représente l'effet du nœud (gène) X^j vers le nœud X^i, sous les hypothèses suivantes :

1. le graphe $\mathcal{G}_0$ associé à la matrice G_0 est un *Graphe Acyclique Dirigé* (DAG),
2. les variables aléatoires $(X^i)_{1\leq i \leq p}$ suivent une loi normale,
3. les variances du bruit sont indépendantes du nœud choisi : $\sigma_j^2 = \sigma_{j'}^2$ pour tout $j \neq j'$ (on les supposera égales à 1).

Modèle et hypothèses

On s'intéresse au modèle suivant :

$$\forall j \in [\![1,p]\!], \quad X^j = \sum_{i=1, i\neq j}^{p} (G_0)_i^j \; X^i + \varepsilon^j,$$

où $\varepsilon^j \sim \mathcal{N}(0, \sigma_j^2)$ et $(G_0)_i^j \neq 0$ représente l'effet du nœud (gène) X^j vers le nœud X^i, sous les hypothèses suivantes :

1. le graphe $\mathcal{G}_0$ associé à la matrice G_0 est un *Graphe Acyclique Dirigé* (DAG),
2. les variables aléatoires $(X^i)_{1 \leq i \leq p}$ suivent une loi normale,
3. les variances du bruit sont indépendantes du nœud choisi : $\sigma_j^2 = \sigma_{j'}^2$ pour tout $j \neq j'$ (on les supposera égales à 1).

$\rightarrow$ permet de montrer l'identifiabilité du modèle (Peters, 2014)

Estimateur du maximum de vraisemblance

- On suppose que l'on observe un échantillon $(X_i^1, \ldots, X_i^p)_{1 \leq i \leq n}$ i.i.d. de taille n du modèle précédent, réécrit sous la forme matricielle simplifiée : $X = XG_0 + \varepsilon$.

- Pour reconstruire le vrai réseau, une méthode classique en statistiques consiste à trouver l'estimateur du maximum de vraisemblance

$$\hat{G} = \underset{G \in \mathcal{G}_{DAG}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} \left((X - XG)_i^j \right)^2 \right\}$$

où $\mathcal{G}_{DAG}$ désigne l'ensemble des graphes acycliques dirigés.

Estimateur du maximum de vraisemblance

- On suppose que l'on observe un échantillon $(X_i^1, \dots, X_i^p)_{1 \leq i \leq n}$ i.i.d. de taille n du modèle précédent, réécrit sous la forme matricielle simplifiée : $X = XG_0 + \varepsilon$.
- Pour reconstruire le vrai réseau, une méthode classique en statistiques consiste à trouver l'estimateur du maximum de vraisemblance **pénalisée :**

$$\hat{G} = \underset{G \in \mathcal{G}_{DAG}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} \left((X - XG)_i^j \right)^2 + \boldsymbol{\lambda}\, \mathbf{pen}(\boldsymbol{G}) \right\}$$

où $\mathcal{G}_{DAG}$ désigne l'ensemble des graphes acycliques dirigés.

Estimateur du maximum de vraisemblance

- On suppose que l'on observe un échantillon $(X_i^1, \ldots, X_i^p)_{1 \leq i \leq n}$ i.i.d. de taille n du modèle précédent, réécrit sous la forme matricielle simplifiée : $X = XG_0 + \varepsilon$.

- Pour reconstruire le vrai réseau, une méthode classique en statistiques consiste à trouver l'estimateur du maximum de vraisemblance **pénalisée :**

$$\hat{G} = \underset{G \in \mathcal{G}_{DAG}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} \left((X - XG)_i^j \right)^2 + \lambda \|G\|_1 \right\},$$

où $\mathcal{G}_{DAG}$ désigne l'ensemble des graphes acycliques dirigés et

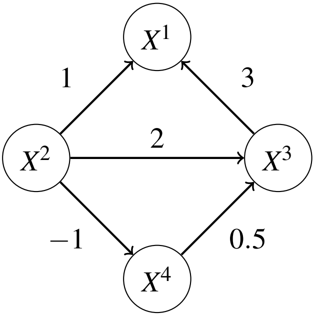
$$\|G\|_1 = \sum_{i,j} \left| G_i^j \right|$$

Décomposition d'un DAG

Une matrice G est compatible avec un DAG $\mathcal{G}$ ssi :

$$G = PT^tP,$$

où P et T sont respectivement une matrice de permutation et une matrice triangulaire inférieure stricte.



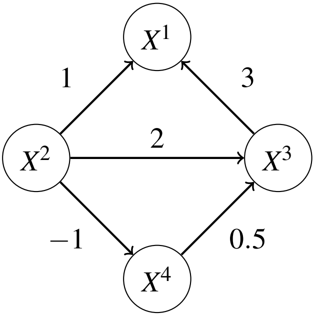
$$G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & -1 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{pmatrix}$$

Décomposition d'un DAG

Une matrice G est compatible avec un DAG $\mathcal{G}$ ssi :

$$G = PT^tP,$$

où P et T sont respectivement une matrice de permutation et une matrice triangulaire inférieure stricte.



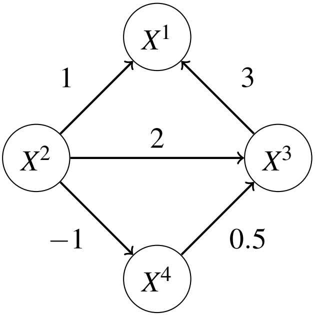
$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 1 & 2 & -1 & 0 \end{pmatrix}$$

Décomposition d'un DAG

Une matrice G est compatible avec un DAG $\mathcal{G}$ ssi :

$$G = PT^tP,$$

où P et T sont respectivement une matrice de permutation et une matrice triangulaire inférieure stricte.



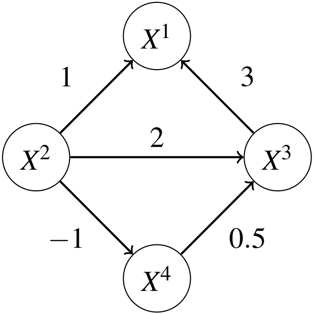
$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Décomposition d'un DAG

Une matrice G est compatible avec un DAG $\mathcal{G}$ ssi :

$$G = PT^tP,$$

où P et T sont respectivement une matrice de permutation et une matrice triangulaire inférieure stricte.



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

$\rightarrow \sigma = (1\ \ 3\ \ 4\ \ 2) \in \mathfrak{S}_p$ permutation qui ordonne les noeuds du graphe.

Décomposition d'un DAG

Une matrice G est compatible avec un DAG $\mathcal{G}$ ssi :

$$G = PT^tP,$$

où P et T sont respectivement une matrice de permutation et une matrice triangulaire inférieure stricte.

On se ramène au problème de minimisation suivant :

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

où $\|X\|_F^2 = \sum_{i,j=1}^p \left(X_i^j\right)^2$.

D'un point de vue théorique

Considérons $\hat{G}(\lambda) = \hat{P}\hat{T}^t\hat{P}$, où

$$(\hat{P}, \hat{T}) = \operatorname*{argmin}_{P \in \mathbb{P}_p(\mathbb{R}), T \in \mathbb{T}_p(\mathbb{R})} \left\{ \frac{1}{n} \|X - XPT^tP\|_F^2 + \lambda \|T\|_1 \right\}.$$

Notons Π_0 l'ensemble des permutations telles que :

$$\Pi_0 = \{P \in \mathbb{P}_p(\mathbb{R}), {}^tPG_0P \text{ est triangulaire inférieure stricte}\}.$$

Deux objectifs :

- Est-ce que $\hat{P}$ appartient à Π_0 ?
- Obtention d'inégalités comparant $\hat{G}(\lambda)$ avec G_0 ?

Hypothèses sur le modèle

$\mathbf{H_{cov}}$ $\max_{1\leq j\leq p} \mathrm{Var}(X^j) \leq \sigma_0^2$, où σ_0^2 est indépendant de p, n.

$\mathbf{H_{RE}(s)}$ Pour $s \in [\![1, p^2]\!]$,

$$\kappa(s) := \min_{\mathcal{S}\subset[\![1,p]\!]^2, |\mathcal{S}|\leq s} \quad \min_{\left\|M_{\mathcal{S}^C}\right\|_1 \leq 3\|M_{\mathcal{S}}\|_1} \frac{\|XM\|_F}{\sqrt{n}\,\|M_{\mathcal{S}}\|_F} > 0.$$

$\mathbf{H_S}$ Le nombre maximal de parents s_{max} d'un nœud du graphe satisfait :

$$s_{max} \leq C\sqrt{\frac{n}{\log p}}p^{-3/2}.$$

Hypothèses sur le modèle

$\mathbf{H_{cov}}$ $\max_{1\leq j\leq p} \mathrm{Var}(X^j) \leq \sigma_0^2$, où σ_0^2 est indépendant de p, n.

$\mathbf{H_{RE}(s)}$ Pour $s \in [\![1, p^2]\!]$,

$$\kappa(s) := \min_{\mathcal{S}\subset[\![1,p]\!]^2, |\mathcal{S}|\leq s} \quad \min_{\left\|M_{\mathcal{S}^C}\right\|_1 \leq 3\|M_{\mathcal{S}}\|_1} \frac{\|XM\|_F}{\sqrt{n}\,\|M_{\mathcal{S}}\|_F} > 0.$$

$\mathbf{H_S}$ Le nombre maximal de parents s_{max} d'un nœud du graphe satisfait :

$$s_{max} \leq C\sqrt{\frac{n}{\log p}}p^{-3/2}.$$

$\mathbf{H_{dim}}$ $p^3 \log p = \underset{n\to+\infty}{\mathcal{O}}(n).$

Hypothèses sur le modèle

$\mathbf{H_{cov}}$ $\max_{1\leq j\leq p} \mathrm{Var}(X^j) \leq \sigma_0^2$, où σ_0^2 est indépendant de p, n.

$\mathbf{H_{RE}(s)}$ Pour $s \in [\![1, p^2]\!]$,

$$\kappa(s) := \min_{\mathcal{S}\subset[\![1,p]\!]^2, |\mathcal{S}|\leq s} \; \min_{\left\|M_{\mathcal{S}^C}\right\|_1 \leq 3\|M_{\mathcal{S}}\|_1} \frac{\|XM\|_F}{\sqrt{n}\,\|M_{\mathcal{S}}\|_F} > 0.$$

$\mathbf{H_S}$ Le nombre maximal de parents s_{max} d'un nœud du graphe satisfait :

$$s_{max} \leq C\sqrt{\frac{n}{\log p}} p^{-3/2} \leq \mathbf{C}.$$

$\mathbf{H_{dim}}$ $p^3 \log p = \underset{n\to+\infty}{\mathcal{O}}(n)$.

D'un point de vue théorique

Considérons $\hat{G}(\lambda)$ avec $\lambda = 2C\sqrt{\frac{\log p}{n} s_{max}}$. Sous les hypothèses $\mathbf{H_{cov}}$, $\mathbf{H_S}$, $\mathbf{H_{dim}}$ et une condition d'identifiabilité*, avec probabilité au moins $1 - 5/p$:

$$\hat{P} \in \Pi_0.$$

** Hypothèse permettant de s'assurer qu'on apprend correctement l'ordre des variables.*

D'un point de vue théorique

Considérons $\hat{G}(\lambda)$ avec $\lambda = 2C\sqrt{\frac{\log p}{n} s_{max}}$. Sous les hypothèses $\mathbf{H_{cov}}$, $\mathbf{H_S}$, $\mathbf{H_{dim}}$ et une condition d'identifiabilité*, avec probabilité au moins $1 - 5/p$:

$$\hat{P} \in \Pi_0.$$

Théorème (Inégalités en prédiction et estimation)

De plus, si on suppose que l'hypothèse $\mathbf{H_{RE}}(\mathbf{s})$ *est satisfaite pour un certain* $1 \leq s \leq p^2$*, alors, avec probabilité au moins* $1 - 5/p$ *:*

$$\frac{1}{n}\left\|X\hat{G} - XG_0\right\|_F^2 \leq \frac{16C^2 s_{max}^2 \log p}{n\kappa^2(s)},$$

$$\left\|\hat{G} - G_0\right\|_1 \leq \frac{16C}{\kappa^2(s)}\sqrt{\frac{\log p}{n}} s_{max}^{3/2}.$$

* *Hypothèse permettant de s'assurer qu'on apprend correctement l'ordre des variables.*

Procédure d'optimisation

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

1. Si on connaît $\hat{P} \in \mathbb{P}_p(\mathbb{R})$, le problème devient convexe et on peut trouver $\hat{T} \in \mathbb{T}_p(\mathbb{R})$ solution de :

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

Procédure d'optimisation

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

1. Si on connaît $\hat{P}\in\mathbb{P}_p(\mathbb{R})$, le problème devient convexe et on peut trouver $\hat{T}\in\mathbb{T}_p(\mathbb{R})$ solution de :

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

2. L'exploration de l'ensemble des permutations est difficile étant données sa taille et sa structure (ensemble discret et non convexe).

Procédure d'optimisation

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

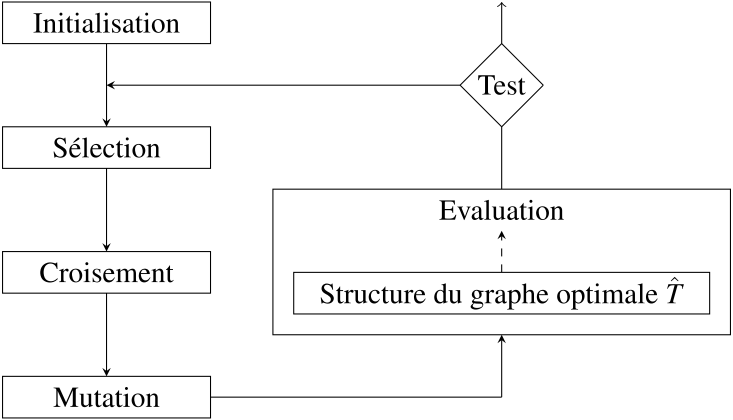
1. Si on connaît $\hat{P} \in \mathbb{P}_p(\mathbb{R})$, le problème devient convexe et on peut trouver $\hat{T} \in \mathbb{T}_p(\mathbb{R})$ solution de :

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

2. L'exploration de l'ensemble des permutations est difficile étant données sa taille et sa structure (ensemble discret et non convexe).

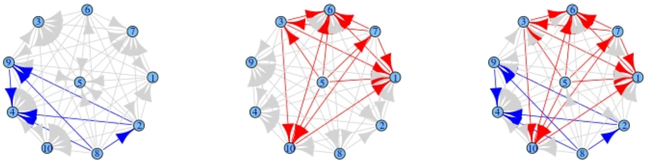
$\rightarrow$ Mise en place d'un algorithme génétique hybride.

Algorithme génétique hybride



Opérateur de croisement

$$\begin{array}{ccccccccccc}
 & 4 & 3 & 10 & 7 & 5 & 9 & 1 & 2 & 6 & 8 \\
+ & 6 & 1 & \cancel{9} & \cancel{4} & 10 & \cancel{2} & \cancel{8} & 3 & 7 & 5 \\
= & 4 & * & * & * & * & 9 & * & 2 & * & 8
\end{array}$$

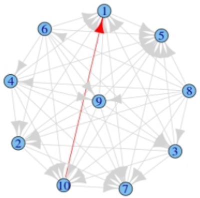


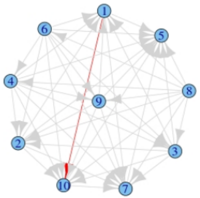
Opérateur de mutation

5	7	1	10	2	3	9	4	6	8

$$\downarrow$$

5	7	10	1	2	3	9	4	6	8





Résultats numériques I

Pour une valeur de paramètre de pénalisation λ donnée, on compare les arêtes apprises avec les arêtes du vrai réseau afin de calculer

- la sensibilité (ou rappel), mesurant le ratio entre arêtes correctement prédites et arêtes à prédire,

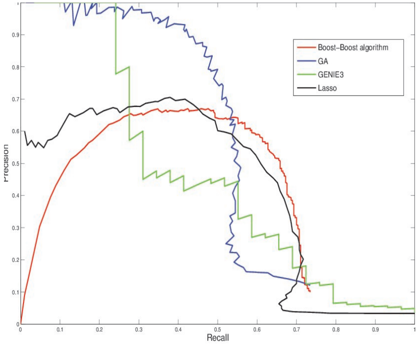
$$\text{sensibilité} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Négatifs}}.$$

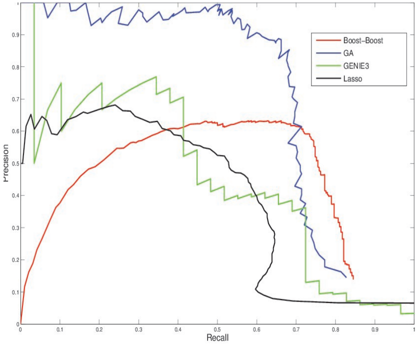
- la précision, mesurant le ratio entre arêtes correctement prédites et arêtes prédites,

$$\text{précision} = \frac{\text{Vrais Positifs}}{\text{Vrais Positifs} + \text{Faux Positifs}},$$

ce qui permet de tracer des courbes précision-rappel en faisant varier le paramètre λ.

Résultats numériques I





Courbes précision-rappel pour deux réseaux en “étoile” de taille 30

algorithme génétique (en bleu)

forêts aléatoires (en vert)

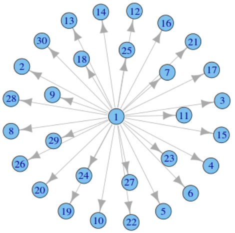
algorithme de Boosting multi-tâches (en rouge)

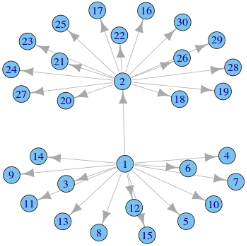
Lasso (en noir).

Résultats numériques II

Méthode	algo génétique	Boosting	Forêts aléatoires	Lasso
N1	**0.5223**	0.3947	0.4212	0.4312
N2	**0.6919**	0.4315	0.4136	0.3920
N3	0.3557	0.3596	0.2718	**0.4112**
N4	**0.4171**	0.2332	0.2887	0.3349

TABLE: Aire sous la courbe précision-recall pour l'ensemble des réseaux et l'état de l'art.





Synthèse des travaux

- Mise en place d'un algorithme pour la résolution de régressions parcimonieuses multi-tâches.

Synthèse des travaux

- Mise en place d'un algorithme pour la résolution de régressions parcimonieuses multi-tâches.
- Adaptation de l'algorithme de $\mathbb{L}_2$-Boosting pour l'estimation d'indices de sensibilité dans le cas de variables dépendantes.

Synthèse des travaux

- Mise en place d'un algorithme pour la résolution de régressions parcimonieuses multi-tâches.
- Adaptation de l'algorithme de $\mathbb{L}_2$-Boosting pour l'estimation d'indices de sensibilité dans le cas de variables dépendantes.
- Apprentissage de relations causales au sein du graphe : adaptation des résultats théoriques obtenus par van de Geer *et al.* (2013) pour une pénalité en norme ℓ_0 et implémentation d'un algorithme génétique hybride.

Conclusion

- Les algorithmes de $\mathbb{L}_2$-Boosting constituent des méthodes de régression performantes :
  - Pas d'hypothèse nécessaire sur la structure de corrélation du design (cas $m = 1$),
  - temps de calcul inférieurs à ceux des méthodes de l'état de l'art,
  - adaptation au cas multi-tâches pour améliorer l'apprentissage de réseaux,
  - particulièrement adapté à l'analyse de sensibilité.

Conclusion

- Les algorithmes de $\mathbb{L}_2$-Boosting constituent des méthodes de régression performantes :
  - Pas d'hypothèse nécessaire sur la structure de corrélation du design (cas $m = 1$),
  - temps de calcul inférieurs à ceux des méthodes de l'état de l'art,
  - adaptation au cas multi-tâches pour améliorer l'apprentissage de réseaux,
  - particulièrement adapté à l'analyse de sensibilité.

- Amélioration de l'apprentissage de réseaux en s'intéressant aux relations causales entre les nœuds du réseau
  - gérer les problèmes d'identifiabilité (restriction au cas où les variances du bruit sont égales),
  - exploration de l'ensemble des DAGs par un algorithme génétique hybride.

Perspectives

- Les algorithmes de $\mathbb{L}_2$-Boosting :
  - Affaiblir les hypothèses $\mathbf{H_{RE^{\pm}}}$ en autorisant $\lambda_{max,min}$ à croître avec n,
  - Affaiblir l'hypothèse $\mathbf{H_b^1}$ en considérant des bases de fonctions plus grandes.

Perspectives

- Les algorithmes de $\mathbb{L}_2$-Boosting :
  - Affaiblir les hypothèses $\mathbf{H}_{\mathbf{RE}^{\pm}}$ en autorisant $\lambda_{max,min}$ à croître avec n,
  - Affaiblir l'hypothèse $\mathbf{H}_{\mathbf{b}}^{\mathbf{1}}$ en considérant des bases de fonctions plus grandes.
- L'algorithme génétique hybride :
  - Le tester sur des jeux de données réels, ou de plus grande dimension,
  - Faire un package R,
  - Montrer des résultats de convergence de l'algorithme génétique.

Perspectives

- Les algorithmes de $\mathbb{L}_2$-Boosting :
    - Affaiblir les hypothèses $\mathbf{H}_{\mathbf{RE}^{\pm}}$ en autorisant $\lambda_{max,min}$ à croître avec n,
    - Affaiblir l'hypothèse $\mathbf{H}^{\mathbf{1}}_{\mathbf{b}}$ en considérant des bases de fonctions plus grandes.
- L'algorithme génétique hybride :
    - Le tester sur des jeux de données réels, ou de plus grande dimension,
    - Faire un package R,
    - Montrer des résultats de convergence de l'algorithme génétique.
- Gérer les problèmes d'identifiabilité du modèle en considérant des données d'intervention pour orienter les arêtes du graphe apprises sans hypothèses supplémentaires sur les variances du bruit.

Merci pour votre attention !

- E. Candes, T. Tao. The Dantzig selector : statistical estimation when p is much larger than n. The Annals of Statistics, 35(6) : 2313-2351, 2007.
- M. Champion, C. Cierco-Ayrolles, S. Gadat, M. Vignes. Sparse regression and support recovery with $\mathbb{L}_2$-Boosting algorithms. JSPI, 2014, à paraître.
- M. Champion, G. Chastaing, S. Gadat, C. Prieur. $\mathbb{L}_2$-Boosting for sensitivity analysis with dependent inputs. Statistica Sinica, 2014, a paraître.
- G. Chastaing, F. Gamboa, C. Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables. Electronic Journal of Statistics, 6 :2420-2448, 2012.
- V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts. Inferring regulatory networks from expression data using tree-based methods. PLoS ONE, 5(9) : e12776, 2010.
- B. Liu, A. de la Fuente, I. Hoeschele. Gene network inference via structural equation modeling in genetical genomics experiments. Genetics, 178 : 1763–1776, 2008.
- J. Peters, P. Bühlmann. Identifiability of Gaussian structural equation models with equal error variances. Biometrika, 101 :219-228, 2014.
- R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, 58(1) : 267-288, 1996.
- S. van de Geer, P. Bühlmann. ℓ_0-penalized maximum likelihood for sparse directed acyclic graphs. The Annals of Statistics,41(2) : 536-567.