

Estimation bayésienne asymptotique de la structure d'un graphe initialisée par Graphical lasso

Aude COSTARD^{1,2}, Sophie ACHARD¹, Olivier MICHEL¹, Pierre BORGNAT², Patrice ABRYS²

¹GIPSA-lab

11 rue des Mathématiques, BP 46, 38402 Saint Martin d'Hères Cedex, France

²Laboratoire de Physique, ENS de Lyon, CNRS UMR 5672,
46 allée d'Italie, 69007 Lyon, France

¹ prénom.nom@gipsa-lab.grenoble-inp.fr, ² prénom.nom@ens-lyon.fr

Résumé – Lors de l'étude de processus dont les observations sont des séries temporelles multivariées, il est intéressant de connaître les relations de dépendance conditionnelle. Pour les séries temporelles gaussiennes, les modèles graphiques gaussiens permettent de représenter ces relations. Dans cet article, nous proposons une stratégie, ABiGlasso (Asymptotic Bayesian method initialised by Graphical lasso), qui estime la structure de graphe représentant les dépendances conditionnelles d'un processus en tirant profit de la rapidité du Graphical lasso et des performances d'une approche bayésienne. Les solutions obtenues sur des simulations dont la structure de dépendance conditionnelle est connue sont très proches, au sens de la distance de Hamming, de la solution attendue. De plus, cette stratégie a un coût de calcul inférieur à celui d'une recherche bayésienne exhaustive.

Abstract – When studying process with multivariate time series, a point of interest is the knowledge about conditional dependency. For Gaussian times series, Gaussian graphical models are commonly used to represent such dependencies. In this paper, we present an approach that estimates the graph structure representing the conditional dependencies of a process. This approach uses Graphical lasso to fasten a Bayesian approach. The obtained solutions for datasets simulated from known graph structures are closed, according to the Hamming distance, to the expected solution. Moreover, our approach has a lower computing cost than a Bayesian exhaustive research.

1 Introduction

Nous nous intéressons ici à des processus (ou des systèmes) sur lesquels sont observées des séries temporelles multivariées. Pour l'étude de tels processus, il est intéressant d'identifier la relation de dépendance entre deux variables tout en prenant en compte toute interaction avec d'autres variables de ce processus. Ceci définit la notion de dépendance conditionnelle. Si le processus est assimilé à un jeu de données multivarié gaussien, sa structure de dépendance conditionnelle peut être représentée à l'aide des modèles graphiques gaussiens (MGG) [6, 12]. Pour un MGG, chaque variable correspond à un nœud du graphe et la présence (respectivement l'absence) d'une arête entre deux nœuds signifie que les deux variables associées à ces nœuds sont dépendantes (respectivement indépendantes) conditionnellement aux autres variables.

De nombreuses approches ont été développées pour estimer les MGG. Ces approches se classent principalement en deux catégories : les approches par sélection de covariance et les approches bayésiennes par sélection de modèles. Elles sont notamment utilisées pour l'analyse de données en neuro-imagerie [3] et en génétique [10]. Parmi les méthodes d'estimation des MGG par sélection de covariance, l'algorithme Graphical lasso [4] a pour avantage de donner rapidement une solution. Cette dernière dépend du choix d'un paramètre de pénalisation généralement fixé empiriquement. Par ailleurs, les méthodes d'estimation des MGG par sélection de modèles utilisent essentiellement des approches bayésiennes. Marrelec [7] propose une formule asymptotique qui permet d'avoir la probabilité a posteriori de toutes les structures de graphe sachant un jeu de données. Ces méthodes s'appuient sur une recherche "exhaus-

sive" sur toutes les structures de graphe possibles, ce qui entraîne des temps de calcul importants sinon rédhibitoires.

Cet article présente une stratégie, nommée ABiGlasso (Asymptotic Bayesian method initialized by Graphical lasso), visant à tirer profit des avantages des deux méthodes citées précédemment. L'idée directrice est de développer une recherche bayésienne locale dans un voisinage d'une solution déterminée par Glasso. Une contribution originale de ce papier est d'introduire à la fois une notion de voisinage dans l'espace des graphes et une méthode permettant de s'affranchir du choix du paramètre de pénalisation du Glasso. Nous illustrons les performances de cette stratégie sur des jeux de données simulés. Les solutions obtenues sont très proches, au sens de la distance de Hamming, de la solution exacte tout en assurant des coûts de calcul inférieurs à une recherche exhaustive.

2 Modèles graphiques

2.1 Modèles graphiques gaussiens

$G = (\mathbb{V}, \mathbb{E})$ désigne un graphe non-dirigé où \mathbb{V} et \mathbb{E} sont respectivement l'ensemble des nœuds et des arêtes de G . \mathbb{V} est de dimension $p = |\mathbb{V}|$.

Les modèles graphiques gaussiens (MGG) sont des graphes dont les nœuds correspondent à des variables aléatoires gaussiennes, notées X_i où $\mathbf{X} = (X_1, \dots, X_p)$ est un vecteur aléatoire gaussien : $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$. μ est le vecteur moyenne de taille p et Σ est la matrice de covariance $p \times p$.

Étant donné $(\mathbf{X}^1, \dots, \mathbf{X}^n)$ un échantillon du vecteur aléatoire \mathbf{X} , la matrice de covariance empirique est définie par : $S = 1/n \sum_{k=1}^n (\mathbf{X}^k - \bar{\mathbf{X}})(\mathbf{X}^k - \bar{\mathbf{X}})^T$ avec $\bar{\mathbf{X}} = 1/n \sum_{k=1}^n \mathbf{X}^k$.

$K = \Sigma^{-1}$ définit la matrice de précision.

Avec les notations précédentes, dans le cadre des MGG, l'indépendance conditionnelle entre deux nœuds s'exprime selon l'équivalence suivante :

$$(i, j) \notin \mathbb{E}, i \neq j \Leftrightarrow X_i \perp X_j | X_{V \setminus \{i, j\}}.$$

Les démonstrations sur les propriétés des MGG sont en particulier détaillées dans Lauritzen(1996) [6].

2.2 Méthodes pour estimer les MGG

2.2.1 Approches par sélection de covariance

Dans le cas gaussien, les notions de dépendance conditionnelle et de corrélation partielle sont confondues. La dépendance conditionnelle est donc quantifiée par la corrélation partielle, version normalisée de la matrice de précision. Si deux variables X_i et X_j sont indépendantes conditionnellement aux autres variables alors :

$$X_i \perp X_j | X_{V \setminus \{i, j\}} \Leftrightarrow K_{i, j} = 0.$$

Donc, dans le cadre de l'estimation des MGG :

$$(i, j) \notin \mathbb{E}, i \neq j \Leftrightarrow K_{i, j} = 0. \quad (1)$$

Dès 1972, Dempster [2] introduit la notion de sélection de covariance en proposant une méthode d'estimation itérative de la matrice de précision : chaque itération consiste à proposer l'insertion d'une nouvelle arête. Plus récemment, les méthodes de sélection de covariance se basent sur l'estimation au sens du maximum de vraisemblance de K [12]. Ces méthodes sont formulées sous forme de problèmes de pénalisation l_0 (pour assurer la parcimonie), relaxés en problèmes de pénalisation l_1 pour assurer la convexité (voir [8], par exemple) :

$$\hat{K} = \operatorname{argmax}_K \log(\det(K)) - \operatorname{tr}(KS) - \lambda \|K\|_1 \quad (2)$$

où λ est le paramètre de pénalisation et $\|\cdot\|_1$ est la norme l_1 d'une matrice.

Cette formulation (2) a permis le développement de nombreux algorithmes rapides, dont le Graphical lasso (Glasso) [4]. Le choix du paramètre λ fixe le degré de parcimonie de \hat{K} , c'est-à-dire son nombre de valeurs nulles. Celui-ci est donc crucial. Cependant, ce paramètre est généralement choisi de façon empirique à l'exception notoire de la méthode proposée par Wang [11] qui développe une approche bayésienne pour l'estimation de λ .

2.2.2 Approches bayésiennes par sélection de modèles

Les approches bayésiennes par sélection de modèles ont pour objectif d'obtenir les probabilités a posteriori des différentes structures de dépendances conditionnelles possibles sachant l'échantillon $(\mathbf{X}^1, \dots, \mathbf{X}^n)$ de \mathbf{X} . La solution correspond au maximum a posteriori. Dawid et Lauritzen [1] ont proposé une modélisation de la densité a priori de la matrice de covariance, sachant la structure de graphe, dans le cas des graphes décomposables. Cette densité est la loi Hyper Inverse Wishart (voir [5], par exemple). Cependant, l'ensemble des solutions est limité aux graphes décomposables. D'autres travaux ont permis de créer des approches parcourant aussi bien les graphes décomposables que les graphes non-décomposables [9] mais le nombre de structures de graphe possible entraîne des coûts de calcul rédhibitoires.

Dans [7], Marrelec obtient la probabilité a posteriori de la structure de graphe sachant $\mathbf{y} = (\mathbf{X}^1, \dots, \mathbf{X}^n)$, en calculant

la vraisemblance asymptotique (quand $n \rightarrow \infty$ et sous réserve de convergence de l'algorithme bayésien) de \mathbf{y} conditionnellement à la structure de graphe. Cette loi de probabilité a posteriori s'écrit sous la forme suivante :

$$P(G|\mathbf{y}) \stackrel{a}{=} \frac{1}{C(\mathbf{y})} \times \frac{\phi_{\mathbf{p}_{\bar{\mathbb{E}}}, cW_{\bar{\mathbb{E}}\bar{\mathbb{E}}}}(\mathbf{0})}{V(G)}. \quad (3)$$

$C(\mathbf{y})$ est une constante qui ne dépend que des observations (remarquons qu'elle ne peut être obtenue que par le calcul de $P(G|\mathbf{y})$ pour tout G). Ci-après, $\bar{\mathbb{E}}$ indexe les arêtes non présentes dans G (les zéros de K). $\phi_{\mathbf{p}_{\bar{\mathbb{E}}}, cW_{\bar{\mathbb{E}}\bar{\mathbb{E}}}}(\mathbf{0})$ est la valeur de la fonction gaussienne de paramètres $\mathbf{p}_{\bar{\mathbb{E}}}$ et $cW_{\bar{\mathbb{E}}\bar{\mathbb{E}}}$ en $\mathbf{0}$. $\mathbf{p}_{\bar{\mathbb{E}}}$ correspond à la matrice de précision empirique vectorisée, c est une constante de mise à l'échelle ($c = 1/(n + p + 1)$) et W est à la matrice d'Isserlis [9] de la matrice de covariance empirique S ; $V(G)$ représente le volume des matrices de précision normalisées associées au graphe G . Cette approche, comme toutes les approches bayésiennes souffre de temps de calcul importants car elle est utilisée pour une recherche exhaustive.

3 ABiGlasso

La méthode ABiGlasso (Asymptotic Bayesian method initialised by Graphical lasso) a pour objectif de tirer profit des avantages du Glasso et d'une approche bayésienne basée sur la formule (3). L'originalité de la méthode est, d'une part, de proposer une recherche bayésienne locale sur un voisinage d'une solution déterminée par Glasso (section 3.1) et, d'autre part, de s'affranchir du choix du λ du Glasso (section 3.2).

3.1 Profil de probabilité pour une valeur de λ

Dans un premier temps, nous fixons la valeur de λ . Le graphe solution proposé par le Glasso est noté G_λ^{Gl} et dépend donc de λ . Pour chaque solution, nous définissons le voisinage du graphe G_λ^{Gl} , noté $\mathcal{G}_\lambda(e)$, comme l'ensemble des graphes qui diffèrent de G_λ^{Gl} d'au plus e arêtes.

Pour chaque graphe G du voisinage $\mathcal{G}_\lambda(e)$, à partir de la formule (3), nous calculons le score $C(\mathbf{y})P(G|\mathbf{y})$. Le graphe le plus probable sur $\mathcal{G}_\lambda(e)$ est donné par le maximum sur l'ensemble des scores obtenus. Ce graphe est noté G_λ^{map} et son score $C(\mathbf{y})P(G_\lambda^{map}|\mathbf{y})$ est noté $z(\lambda|\mathbf{y})$.

L'utilisation de la formule (3) est coûteuse en temps de calcul car il faut estimer le volume $V(G)$ pour chaque graphe exploré. En contraignant la recherche à un voisinage d'une solution du Glasso, nous diminuons le temps de calcul tout en conservant l'avantage des approches bayésiennes de fournir une probabilité a posteriori pour les graphes explorés. Le gain en temps de calcul reste important malgré le coût de l'estimation de $V(G)$ (les temps de calcul sur différents exemples sont reportés dans le Tableau 1). Cette estimation limite actuellement l'applicabilité de l'approche proposée à des graphes à $|\mathbb{V}| \leq 8$.

3.2 Ensemble de λ parcouru

Nous avons montré comment obtenir le graphe G_λ^{map} de plus fort score $z(\lambda|\mathbf{y})$ pour une valeur de λ en exploitant la formule (3). Nous proposons de faire varier λ sur un ensemble fini Λ . En appliquant $|\Lambda|$ fois l'algorithme Glasso, autant d'ensembles $\mathcal{G}_\lambda(e)$ sont construits. La définition des scores à partir de l'équation (3) peut s'appliquer à chacun de ces ensembles. Nous notons $\lambda^* = \operatorname{argmax}_\lambda z(\lambda|\mathbf{y})$. Étant donné qu'un même graphe peut appartenir à plusieurs voisinages $\mathcal{G}_\lambda(e)$, il n'y a pas unicité de λ^* .

Le coût de calcul augmente au plus linéairement en fonction de $|\Lambda|$ (les valeurs de $V(G)$ déjà calculées pouvant être conservées). Plus le nombre de valeurs de λ parcourues est grand, plus $G_{\lambda^*}^{map}$ peut se rapprocher de la solution exacte au sens de la distance de Hamming conduisant à un compromis coût/performance.

Dans la suite, Λ est obtenu par discrétisation d'un intervalle $[\lambda^{min}, \lambda^{max}]$ en utilisant une recherche dichotomique.

La recherche dichotomique est appliquée à la fonction distance $d(\lambda) = \mathbb{E}_{\lambda}^{Gl} - \mathbb{E}_{\lambda}^{map}$ en cherchant parallèlement la borne inférieure λ^{min} sur $[l_g^{min}, l_d^{min}]$ et la borne supérieure λ^{max} sur $[l_g^{max}, l_d^{max}]$. Les intervalles $[\lambda^{min}, \lambda^{max}]$, $[l_g^{min}, l_d^{min}]$ et $[l_g^{max}, l_d^{max}]$ ont pour valeurs initiales l'intervalle $[0, 10]$. L'algorithme converge en temps fini vers l'intervalle $[\lambda_{\infty}^{min}, \lambda_{\infty}^{max}]$. En effet, nous avons observé sur nos résultats de simulations que au moins une valeur de λ^* se trouve dans le plus grand intervalle pour lequel $\text{abs}(|\mathbb{E}_{\lambda}^{Gl}| - |\mathbb{E}_{\lambda}^{map}|) < e$. Nous montrons que cet intervalle est fini car plus λ est grand, plus le nombre d'arêtes diminue.

La figure 1 illustre le procédé dichotomique sur la borne inférieure pour un exemple synthétique où la solution recherchée est $\lambda_{\infty}^{min} = 7$, pour $e = 3$. A chaque itération, λ^{min} est mis à jour : $\lambda^{min}(i) = \frac{l_g^{min}(i) + l_d^{min}(i)}{2}$. La vignette 1.A illustre la convergence de λ^{min} pour un succession de pas de la recherche dichotomique. En parallèle, la vignette 1.B représente les variations de $d(\lambda^{min})$ pour ces mêmes itérations. A la première itération, $d(\lambda^{min}(1)) = 3$, donc $\lambda^{min}(1) < \lambda_{\infty}^{min}$, cela implique la mise à jour suivante : $l_d^{min}(2) = \lambda^{min}(1)$. A l'itération 2, $d(\lambda^{min}(2)) = -2$, et donc $\lambda^{min}(2) > \lambda_{\infty}^{min}$, cela implique la mise à jour suivante : $l_g^{min}(3) = \lambda^{min}(2)$. Le principe est le même jusqu'à ce que la convergence soit atteinte, quand $l_d - l_g < 0.02$.

Si $d(\lambda^{max}(i)) = e$, les valeurs $l_g^{min}(i+1)$ et $l_g^{max}(i+1)$ valent $\lambda^{max}(i)$ et les valeurs $l_d^{min}(i+1)$ et $l_d^{max}(i+1)$ valent $10\lambda^{max}(i)$. Ce test permet, le cas échéant, de s'affranchir de l'arbitraire de l'intervalle initial $[0, 10]$.

Fig 1.A

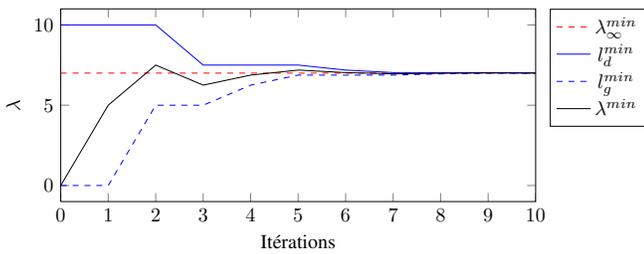


Fig 1.B

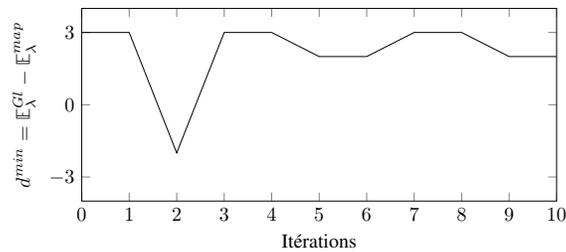


FIGURE 1 – Exemple du comportement du procédé dichotomique pour trouver l'intervalle optimal pour λ sur la borne inférieure λ^{min} pour $\lambda_{\infty}^{min} = 7$.

4 Résultats

Les performances d'ABiGlasso ont été testées sur des jeux de données simulés. Nous nous sommes concentrés sur des graphes à six nœuds car ceux-ci permettent d'avoir des structures diverses (à la différence des graphes à quatre nœuds) tout en assurant des temps de calcul de quelques dizaines de minutes.

4.1 Présentations des jeux de données simulés utilisés

La figure 2 présente les quatre structures de graphe que nous avons choisies d'étudier : 2a est non-décomposable et 2b, 2c et 2d sont décomposables. De plus 2c est séparable et 2d présente un cluster. Pour chaque structure de graphe, notre but est de simuler des matrices de précision respectant la propriété (1) des MGG, c'est-à-dire ayant des valeurs nulles là où il n'y pas d'arêtes dans la structure. Pour simuler les matrices de pré-

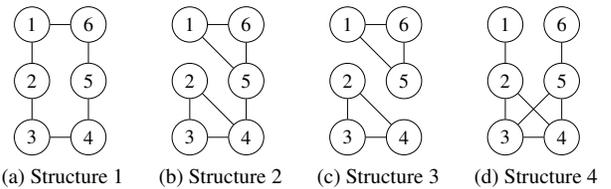


FIGURE 2 – Différentes structures de graphe étudiées

sion, il faut faire face à deux contraintes. D'une part, les matrices doivent être définies positives : la structure de graphe étant fixée, seules les entrées correspondant aux arêtes peuvent être réglées pour respecter cette propriété. D'autre part, afin de limiter le taux de faux négatif, c'est-à-dire le nombre d'arêtes non détectées, nous avons contraint les entrées non nulles à avoir une valeur supérieure ou égale à 0.45. Ces deux contraintes rendent difficile la simulation des matrices de précision.

Les jeux de données simulés sont des échantillons multivariés gaussiens de 600 observations, de moyenne nulle et dont la matrice de covariance est l'inverse de la matrice de précision parcimonieuse générée précédemment.

Les performances sur chaque configuration sont évaluées à l'aide de la distance de Hamming entre le graphe obtenu et le graphe attendu : cette distance mesure le nombre d'arêtes qui diffèrent entre les deux graphes. Concernant le temps de calcul, il est mesuré sur un processeur 2.66 GHz Intel Core i7.

4.2 Résultats

L'algorithme ABiGlasso est appliqué sur les jeux de données simulés à partir des quatre structures de graphe présentées figure 2.

Pour pouvoir comparer les performances des différents jeux de données simulés, l'intervalle optimisé pour λ est normalisé afin d'avoir le même pas entre deux valeurs quelque soit le jeu de données considéré : $\lambda \in [\lambda_a^{min}, 0.2, \lambda_a^{max}]$ avec λ_a^{min} et λ_a^{max} des multiples de 0.2 tels que $[\lambda_{\infty}^{min}, \lambda_{\infty}^{max}] \subset [\lambda_a^{min}, \lambda_a^{max}]$ au plus juste.

Le comportement de l'algorithme est observé pour différentes valeurs du paramètre de voisinage. Dans la suite, nous dissociions le paramètre de voisinage utilisé pour la recherche de l'intervalle optimisé sur λ , que nous notons e_{RI} , et le paramètre utilisé lors du parcours des $\lambda \in [\lambda_a^{min}, \lambda_a^{max}]$, que nous notons e_{main} .

Pour le jeu de données correspondant à la structure (2d),

	e_{RI}	$e_{main} = 1$			$e_{main} = 2$			$e_{main} = 3$			tous les graphes
		1	2	3	1	2	3	1	2	3	
Jeu 1	TC	2m	14m	1h02	10m	14m	1h02	40m	44m	1h14	14h
	DH	2	0	0	1	0	0	0	0	0	0
Jeu 2	TC	3m	20m	1h27	8m	21m	1h27	28m	44m	1h22	14h
	DH	2	2	2	1	1	1	0	0	0	0
Jeu 3	TC	3m	19m	1h19	8m	19m	1h19	28m	45m	1h19	14h
	DH	0	0	0	0	0	0	0	0	0	0
Jeu 4	TC	2m	15m	1h19	6m	15m	1h19	24m	37m	1h26	14h
	DH	4	4	4	3	3	2	2	2	1	0

TC : temps de calcul, DH : distance de Hamming

TABLEAU 1 – Temps de calcul et distances de Hamming pour différentes valeurs des voisinages e_{RI} et e_{main} . $e_{RI} = x$ correspond à ABiGlasso sur un intervalle de λ arbitraire. La colonne tous les graphes correspond à une recherche exhaustive en utilisant la formule (3). En gris, la combinaison de paramètre de voisinage optimale pour le compromis entre temps de calcul et distance de Hamming.

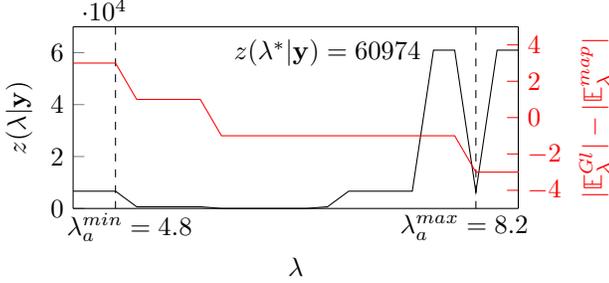


FIGURE 3 – Résultats sur le jeu de données correspondant à la structure (2d). En noir le score de G_λ^{map} , en fonction de λ , obtenu pour $e_{RI} = 3$ et $e_{main} = 3$. En rouge, l'évolution de la différence d'arêtes entre G_λ^{GI} et G_λ^{map} en fonction de λ et en pointillé l'intervalle optimisé sur λ .

la figure 3 illustre le comportement d'ABiGlasso pour les paramètres $e_{RI} = e_{main} = 3$. Dans cet exemple, nous obtenons comme solution l'intervalle $[\lambda_a^{min}, \lambda_a^{max}] = [4.8, 8.2]$. La courbe noire représente le score de G_λ^{map} en fonction de λ . Dans cet exemple, λ^* prend deux valeurs : 7.8 et 8.

Le tableau 1 résume les performances d'ABiGlasso sur les jeux de données simulés à partir des quatre structures de graphe présentées figure 2 pour différentes valeurs de e_{RI} et de e_{main} . La colonne tous les graphes contient les performances obtenues quand $P(G|y)$ est calculé pour toutes les structures de graphe à six nœuds possibles, à partir de la formule (3). Afin de minimiser l'erreur entre la solution obtenue et la structure de graphe attendue, il faut minimiser la distance de Hamming. Pour une valeur de e_{RI} donnée, $e_{main} = 3$ est le paramètre qui minimise cette distance. Concernant le choix de e_{RI} , il faut faire un compromis entre temps de calcul et performances en terme de distance de Hamming. Suite à des simulations sur d'autres jeux de données non présentés ici par soucis de concision, $e_{RI} = 2$ est le meilleur candidat pour équilibrer ce compromis.

En comparaison à une recherche exhaustive, c'est-à-dire au calcul de $P(G|y)$ pour toutes les structures de graphe à six nœuds possibles, l'algorithme ABiGlasso, avec pour paramètres de voisinage $e_{RI} = 2$ et $e_{main} = 3$, a un temps de calcul inférieur à 1h contre 14h pour la recherche exhaustive.

5 Conclusion

Pour des données simulées, l'algorithme ABiGlasso permet d'obtenir une structure de graphe ayant, au plus, deux arêtes de différence avec la structure attendue. De plus, cette méthode s'affranchit du choix du paramètre de pénalisation du Glasso, qui est difficilement contrôlable tout en nécessitant un coût de calcul moindre qu'une recherche exhaustive. Utilisant une approche bayésienne, en plus de donner le graphe optimal G^* et

son paramètre de pénalisation associé, la méthode fournit un profil proportionnel au profil de probabilité sur l'ensemble des graphes parcourus. Ce profil permet de juger de la pertinence de la solution obtenue.

Références

- [1] P. Dawid et S.L. Lauritzen. "Hyper Markov laws in the statistical analysis of decomposable graphical models", *The Annals of Statistics*, vol. 21, pp. 1272-1371, 1993.
- [2] A.P. Dempster. "Covariance selection", *Biometrics*, vol. 28, pp. 157-175, 1972.
- [3] M. Fiecas *et al.* "Quantifying temporal correlations : A test-retest evaluation of functional connectivity in resting-state fMRI", *NeuroImage*, vol. 65, pp.231-241, 2013.
- [4] J. Friedman, T. Hastie et R. Tibshirani. "Sparse inverse covariance estimation with the graphical lasso", *Biostatistics*, vol. 9, pp.432-441, 2007.
- [5] P. Guidici et P.J. Green. "Decomposable graphical gaussian model determination", *Biometrika*, vol.86, pp. 785-801, 1999.
- [6] S.L. Lauritzen. *Graphical Models*, Oxford University press, 1996.
- [7] G. Marrelec et H. Benali. "Asymptotic Bayesian structure learning using graph supports for Gaussian graphical models", *Journal of Multivariate Analysis*, vol. 97, pp. 1451-1466, 2006.
- [8] N. Meinshausen et P. Bühlmann. "High-dimensional graphs and variable selection with the lasso", *The Annals of Statistics*, vol. 34, pp. 1435-1462, 2006.
- [9] A. Roverato. "Hyper Inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian Graphical Models", *Scandinavian Journal of Statistics*, vol. 29, pp. 391-411, 2002.
- [10] T. Ueda *et al.* "Gene expression profiling of hepatitis B- and hepatitis C-related hepatocellular carcinoma using graphical Gaussian modeling", *Genomics*, vol. 101, pp. 238-248, 2013
- [11] H. Wang. "Bayesian Graphical lasso models and efficient posterior computation", *Bayesian Analysis*, vol. 7, pp. 771-790, 2012.
- [12] J. Whittaker. *Graphical models in applied multivariate statistics*, Wiley, 1990.