

Autorégressifs à coefficients variables - Modèles graphiques partiels - Applications aux sciences du vivant

Frédéric Proïa

▶ **To cite this version:**

Frédéric Proïa. Autorégressifs à coefficients variables - Modèles graphiques partiels - Applications aux sciences du vivant. Statistiques [math.ST]. Université d'Angers (UA), 2022. tel-03715444

HAL Id: tel-03715444

https://univ-angers.hal.science/tel-03715444

Submitted on 6 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Habilitation à Diriger des Recherches

Université d'Angers

préparée au Laboratoire Angevin de REcherche en MAthématiques
(LAREMA) – Unité Mixte de Recherche 6093 CNRS

par Frédéric PROÏA

Autorégressifs à coefficients variables – Modèles graphiques partiels – Applications aux sciences du vivant

Habilitation soutenue le 05/07/2022
devant un jury composé de

Gamboa	Fabrice	(Rapporteur)
Gégout-Petit	Anne	(Rapportrice)
Trapani	Lorenzo	(Rapporteur)
Liquet	Benoît	(Examinateur)
Kleptsyna	Marina	(Examinatrice)
Panloup	Fabien	(Directeur de recherche)

Table des matières

Remerciements **3**

Activités de recherche **4**

Introduction **6**

1 Autorégressifs à coefficients variables **17**
1.1 Processus quasi-instables 17
1.2 Processus à coefficients aléatoires 44

2 Modèles graphiques partiels **76**
2.1 Approche par vraisemblance pénalisée 78
2.2 Approche bayésienne 108

3 Applications aux sciences du vivant **142**
3.1 Modélisation de courbes de floraison 142
3.2 Reconstruction probabiliste de généalogies 145

Remerciements

En tout premier lieu, j'adresse mes plus vifs remerciements à Fabrice Gamboa, Anne Gégout-Petit et Lorenzo Trapani qui ont accepté de rapporter ce mémoire, et d'en faire une lecture attentive. Je tiendrai compte très volontairement de tous vos conseils avisés, et je suis par ailleurs honoré de la présence de Benoît Liquet et de Marina Kleptsyna comme examinateurs dans ce jury. Je vous remercie tous très chaleureusement.

Je tiens aussi à avoir une pensée enthousiaste pour mes co-auteurs depuis le début de l'aventure angevine. Tous sont cités dans ce manuscrit, mais je pense tout particulièrement à Fabien Panloup (également directeur de recherche pour cette habilitation), Eunice Okome Obiang, Marius Soltane et Jérémy Clotault qui ont participé de très près à de nombreux travaux dans lesquels je me suis impliqué. Je n'oublie évidemment pas Bernard Bercu, qui m'a lancé dans le monde de la recherche.

Et puis, bien sûr, j'exprime toute ma gratitude à toi, Laure, qui m'accompagnes et me supportes au jour le jour. Ton soutien m'est vraiment indispensable ! Je voudrais pour conclure avoir une pensée forte pour François Ducrot, notre collègue et ami qui vient de nous quitter et que j'estimais profondément.

Activités de recherche

Cette section contient en particulier un récapitulatif de mes thématiques de recherche, la liste de mes publications [1] (à jour en date du 01/07/2022) ainsi qu'un résumé de mes encadrements doctoraux et postdoctoraux. Les articles [1], [2], [4], [5] et [7] seront explicitement mis en avant dans ce manuscrit, les autres seront simplement résumés voire même passés sous silence lorsque trop éloignés des axes directeurs retenus.

Thématiques de recherche privilégiées

— Séries chronologiques et statistique des processus.
— Statistique mathématique.
— Statistique en grande dimension.
— Probabilités et statistiques appliquées.

Publications

[1] A Bayesian approach for partial Gaussian graphical models with sparsity. E. Okome Obiang, P. Jézéquel, F. Proïa. To appear in *Bayesian Analysis.* 2022.

[2] A partial graphical model with a structural prior on the direct links between predictors and responses. E. Okome Obiang, P. Jézéquel, F. Proïa. *ESAIM : Probability and Statistics.* Vol. 25, pp 298-324, 2021.

[3] Comments on the presence of serial correlation in the random coefficients of an autoregressive process. F. Proïa, M. Soltane. *Statistics & Probability Letters.* Vol. 170, 2021.

[4] Moderate deviations in a class of stable but nearly unstable processes. F. Proïa. *Journal of Statistical Planning and Inference.* Vol. 208, pp 66-81, 2020.

[5] Probabilistic reconstruction of genealogies for polyploid plant species. F. Proïa, F. Panloup, C. Trabelsi, J. Clotault. *Journal of Theoretical Biology.* Vol. 462, pp 537-551, 2019.

[6] On the Bickel-Rosenblatt test of goodness-of-fit for the residuals of autoregressive processes. A. Lagnoux, T.M.N. Nguyen, F. Proïa. *ESAIM : Probability and Statistics.* Vol. 23, pp 464-491, 2019.

[7] A test of correlation in the random coefficients of an autoregressive process. F. Proïa, M. Soltane. *Mathematical Methods of Statistics.* Vol. 27 (2), pp 119-144, 2018.

1. Tous les preprints sont disponibles sur ma page personnelle : `http://blog.univ-angers.fr/fredericproia/recherche/`

[8] Testing for residual correlation of any order in the autoregressive process. F. Proïa. *Communications in Statistics – Theory and Methods.* Vol. 47 (3), pp 628-654, 2018.

[9] Stationarity against integration in the autoregressive process with polynomial trend. F. Proïa. *Probability and Mathematical Statistics.* Vol. 38 (1), pp 1-26, 2018.

[10] On the characterization of flowering curves using Gaussian mixture models. F. Proïa, A. Pernet, T. Thouroude, G. Michel, J. Clotault. *Journal of Theoretical Biology.* Vol. 402, pp 75-88, 2016.

[11] On Ornstein-Uhlenbeck driven by Ornstein-Uhlenbeck processes. B. Bercu, F. Proïa, N. Savy. *Statistics & Probability Letters.* Vol. 85, pp 36-44, 2014.

[12] Moderate deviations of functional of Markov processes. V. Bitseki Penda, H. Djellout, L. Dumaz, F. Merlevède, F. Proïa. *ESAIM : Proceedings.* Vol. 44, pp 214-238, 2014.

[13] Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. V. Bitseki Penda, H. Djellout, F. Proïa. *ESAIM : Probability and Statistics.* Vol. 18 (1), pp 308-331, 2014.

[14] A SARIMAX coupled modelling applied to individual load curves intraday forecasting. S. Bercu, F. Proïa. *Journal of Applied Statistics.* Vol. 40 (6), pp 1333-1348, 2013.

[15] Further results on the H-Test of Durbin for stable autoregressive processes. F. Proïa. *Journal of Multivariate Analysis.* Vol. 118, pp 77-101, 2013.

[16] A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. B. Bercu, F. Proïa. *ESAIM : Probability and Statistics.* Vol. 17 (1), pp 500-530, 2013.

Thèse

Thèse débutée en 2010 sous la direction de B. Bercu et soutenue le 04/11/2013 à l'IMB (Bordeaux). [Lien]
Autocorrélation et stationnarité dans le processus autorégressif.

Encadrements

— Co-encadrement (à venir) de Marie Badreau : thèse débutée en 2022 sous la direction d'A. Brouste au LMM (Le Mans).
— Co-encadrement d'Eunice Okome Obiang : thèse débutée en 2019 sous la direction de L. Chaumont au LAREMA, en collaboration avec P. Jézéquel de l'ICO.
Statistique en grande dimension appliquée à la modélisation du cancer du sein métastatique.
— Co-encadrement de Marius Soltane : thèse débutée en 2017 sous la direction d'A. Brouste et soutenue le 10/11/2020 au LMM (Le Mans). [Lien]
Asymptotique de différents processus discrets et continus.
— Co-encadrement de Chiraz Trabelsi : postdoctorat sur la période 2018/2019 au LAREMA, en collaboration avec J. Clotault de l'IRHS.
Reconstruction probabiliste de généalogies.

Introduction

Nous allons résumer dans ce mémoire les principaux travaux réalisés ces dernières années, que l'on va répartir selon trois axes majeurs et quasiment indépendants qui formeront autant de chapitres. En somme, il a été décidé de mettre en avant cinq articles récents dont le contenu explicite est fourni afin d'illustrer les chapitres en proportion globale des travaux présentés, bien que d'autres articles soient également cités. Chaque article donne lieu à un résumé qui n'a pas vocation à servir de recherche bibliographique (nous renvoyons le lecteur au contenu détaillé pour cela), mais simplement de survol des thématiques abordées et des résultats obtenus, afin de se faire en quelques lignes une idée claire de ce qui a été produit. Nous concluons nos résumés par quelques perspectives, pour replacer l'étude dans un contexte plus large et envisager des pistes de développements ultérieurs qui nous semblent particulièrement intéressants. Précisons enfin que les notations utilisées ne correspondent pas toujours à celles des articles en question, dans un souci d'harmonisation du manuscrit. Les notations mathématiques quant à elles paraissent suffisamment usuelles pour ne pas être systématiquement redéfinies.

Résumé des travaux de thèse

Il semble tout d'abord opportun de rappeler en quelques lignes le contenu de ma thèse afin de situer précisément les travaux du Chapitre 1, qui sont en lien direct. Cette dernière se proposait d'étudier le comportement asymptotique de différentes statistiques à la base de tests de corrélation résiduelle et de stationnarité dans une modélisation autorégressive. Rappelons qu'un processus AR(p) centré satisfait l'équation de récurrence

$$\forall\, n \geqslant 1, \quad X_n = \theta^T \Phi_{n-1} + \varepsilon_n \tag{1}$$

où $\theta = (\theta_1, \ldots, \theta_p)^T$ et $\Phi_n = (X_n, \ldots, X_{n-p+1})^T$ pour tout $n \geqslant 0$. Dans la définition usuelle d'une modélisation AR(p), le processus $(\varepsilon_n)_{n \geqslant 1}$ est un bruit blanc, c'est-à-dire une suite de v.a.r. décorrélées, d'espérance nulle et de variance finie $\sigma^2 > 0$. De manière condensée, l'écriture VAR$_p$(1) donnée par

$$\forall\, n \geqslant 1, \quad \Phi_n = C_\theta\, \Phi_{n-1} + E_n \tag{2}$$

est équivalente à la précédente, pour un bruit p-vectoriel $E_n = (\varepsilon_n, 0, \ldots, 0)^T$ et une matrice compagnon donnée par

$$C_\theta = \begin{pmatrix} \theta_1 & \theta_2 & \ldots & \theta_p \\ \ddots & & & \vdots \\ & I_{p-1} & & 0 \\ & & \ddots & \vdots \end{pmatrix}. \tag{3}$$

Il est bien connu que la stabilité du processus p-vectoriel $(\Phi_n)_{n\geqslant 0}$ dépend directement des valeurs propres de C_θ,

$$\rho(C_\theta) = |\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_p|. \tag{4}$$

En particulier, selon (Duflo, 1997, Def. 2.3.17), on dira que le processus est *stable* lorsque $|\lambda_1| < 1$, qu'il est *instable* lorsque $|\lambda_1| = 1$ (*purement instable* si de plus $|\lambda_p| = 1$), ou encore qu'il est *explosif* lorsque $|\lambda_p| > 1$, sans même évoquer tous les cas mixtes qui en découlent. Par ailleurs, puisque

$$\forall\, \lambda \in \mathbb{C}^*, \quad \det(C_\theta - \lambda\, I_p) = (-\lambda)^p\, \theta(\lambda^{-1}) \quad \text{où} \quad \theta(z) = 1 - \theta_1 z - \ldots - \theta_p\, z^p \tag{5}$$

est le polynôme autorégressif du processus, il est clair que toute valeur propre de C_θ est l'inverse d'une racine de θ, de sorte que la condition de stabilité $|\lambda_1| < 1$ est équivalente à l'hypothèse que le polynôme θ est causal, c'est-à-dire que toutes ses racines sont à l'extérieur strict du disque unité $\mathcal{D} = \{z \in \mathbb{C},\, |z| \leqslant 1\}$. Par la suite on pourra avoir recours à la notation $\theta(B)$ pour alléger les écritures autorégressives, avec B l'opérateur retard : $\forall\, t,\, BX_t = X_{t-1}$. Lorsque le processus est défini sur $I \cup \mathbb{N}$ où I contient les indices des valeurs initiales, stabilité et stationnarité (asymptotique) coïncident sous des hypothèses adéquates de moments.

Corrélation résiduelle

Sur une trajectoire observable $(\Phi_0, X_1, \ldots, X_n)$, l'estimateur des moindres carrés de θ s'exprime par

$$\widehat{\theta}_n = \left(\sum_{t=1}^{n} \Phi_{t-1} \Phi_{t-1}^T \right)^{-1} \sum_{t=1}^{n} \Phi_{t-1}\, X_t. \tag{6}$$

Il est usuel d'ajouter une pénalisation de type *ridge* à la matrice de covariance, pour s'exempter d'une hypothèse supplémentaire d'inversibilité, sans pour autant modifier les propriétés asymptotiques de l'estimation. Ces dernière ont aussi été largement étudiées dans la littérature, selon la nature de la perturbation $(\varepsilon_n)_{n\geqslant 1}$, le vecteur initial Φ_0 et les valeurs propres de C_θ. Par exemple dans le cas stable, puisque c'est celui qui nous intéresse dans l'immédiat, avec un bruit blanc fort (i.i.d.) de variance $\sigma^2 > 0$ comme perturbation et Φ_0 garantissant l'ergodicité et la stationnarité stricte du processus (de même loi que Φ_n pour $n \geqslant 1$), l'estimation est fortement consistante et asymptotiquement normale, voir par exemple Brockwell et Davis (2006). Sortant de ce cadre idéal, on pourra relâcher les hypothèses sur le bruit et sur le vecteur initial au prix de quelques hypothèses de moments supplémentaires. Comme indiqué précédemment, pour un Φ_0 arbitraire sur lequel on placera des propriétés de moments utiles à l'étude, le cas stable est assimilable au cas asymptotiquement stationnaire. De même, on sait grâce à Lai et Wei (1983) et Chan et Wei (1988) que la consistance forte et la normalité asymptotique restent vraies lorsque $(\varepsilon_n)_{n\geqslant 1}$ est une différence de $\mathcal{F}_n$-martingale ayant un moment d'ordre $2+\epsilon$ ($\epsilon > 0$), où $\mathcal{F}_n = \sigma(\Phi_0, (\varepsilon_t,\, 1 \leqslant t \leqslant n))$. La nature du bruit est une donnée primordiale dans la manière d'aborder les démonstrations, et d'un point de vue pratique il est intéressant de connaître en détail les conséquences inférentielles de l'introduction de corrélation dans la perturbation. Précisément, le cas où $(\varepsilon_n)_{n\geqslant 1}$ se comporte comme un AR(1) est à l'origine

du test de Durbin-Watson. En particulier, sous des conditions impliquant la stabilité dans le modèle

$$\forall\, n \geqslant 1, \quad \left\{ \begin{array}{rcl} X_n & = & \theta^T \Phi_{n-1} + \varepsilon_n \\ \varepsilon_n & = & \rho\, \varepsilon_{n-1} + V_n \end{array} \right. \tag{7}$$

où la perturbation $(V_n)_{n \geqslant 1}$ est un bruit blanc fort de moment d'ordre 4 fini, on montre dans Proïa (2013), et dans Bercu et Proïa (2013) lorsque $p = 1$, qu'il existe une valeur limite θ^* $(\neq \theta)$ et une matrice Σ_θ définie positive telles que

$$\lim_{n \to +\infty} \widehat{\theta}_n = \theta^* \ \text{ p.s.} \quad \text{et} \quad \sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\theta). \tag{8}$$

Cela signifie que l'estimation perd sa propriété de consistance lorsqu'il existe de la corrélation dans le bruit. Dans un tel contexte, il est assez répandu en économétrie notamment d'utiliser la statistique de Durbin-Watson, voir Durbin et Watson (1950, 1951, 1971), définie à l'aide des résidus issus de l'estimation de θ par

$$\widehat{D}_n = \frac{\sum_{t=1}^n \left(\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-1}\right)^2}{\sum_{t=1}^n \widehat{\varepsilon}_t^2} \tag{9}$$

pour un test de corrélation résiduelle de $\mathcal{H}_0$: "$\rho = 0$" contre $\mathcal{H}_1$: "$\rho \neq 0$", ce que nos travaux permettent de formaliser. Citons également Bitseki Penda *et al.* (2014), travail dans lequel nous établissons pour $p = 1$ des principes de déviations modérées sous des hypothèses spécifiques, parmi lesquels

$$\lim_{n \to +\infty} \frac{1}{b_n^2} \ln \mathbb{P}\left(\frac{\sqrt{n}}{b_n}\big(\widehat{D}_n - D^*\big) \in H\right) = -\inf_{x \in H} I_D(x) \tag{10}$$

pour toute vitesse $1 \ll b_n \ll \sqrt{n}$ et tout $H \in \mathcal{B}(\mathbb{R})$ tel que l'infimum de I_D sur H° et celui sur $\bar{H}$ coïncident, ainsi qu'une valeur limite D^* et une fonction de taux I_D bien identifiées. Certains auteurs comme Samoura (2017) ont par la suite étendu nos résultats au cas où $p \geqslant 1$. Enfin, nous proposons dans Bercu *et al.* (2014) une version continue du modèle avec pour motivation principale la similitude entre l'autorégression d'ordre 1 et le processus d'Ornstein-Uhlenbeck. Il est intéressant de constater que les estimateurs convergent à la même vitesse mais que les valeurs limites sont différentes. Tous ces travaux reposent essentiellement sur la théorie des martingales.

Stationnarité

Les procédures de Dickey-Fuller pour $p = 1$ et de Dickey-Fuller augmenté (ADF) pour $p \geqslant 1$ sont couramment utilisées pour tester la présence d'une racine unitaire dans la modélisation d'une trajectoire supposée AR(p). Ces dernières reposent sur un test de significativité dans un modèle de régression bien particulier. En cas de rejet, on pourra retenir la stationnarité de la série, éventuellement autour d'une tendance constante ou linéaire. Afin de tester la stationnarité et donc d'obtenir une procédure statistique complémentaire, Leybourne et McCabe (1994) introduisent une écriture AR(p) causale modifiée que l'on peut résumer par

$$\forall\, n \geqslant 1, \quad \theta(B)\, X_n = T_n + S_n^\eta + \varepsilon_n \tag{11}$$

où $(T_n)_{n\geqslant 1}$ est une tendance linéaire ou constante, $(\varepsilon_n)_{n\geqslant 1}$ est un bruit blanc de variance $\sigma_\varepsilon^2 > 0$ et $(S_n^\eta)_{n\geqslant 0}$ est une marche aléatoire démarrant en $S_0^\eta = 0$, engendrée par un bruit blanc $(\eta_n)_{n\geqslant 1}$ de variance $\sigma_\eta^2 \geqslant 0$ et indépendant du précédent. Le test de Leybourne-McCabe (LMC) se ramène alors au test de $\mathcal{H}_0$: "$\sigma_\eta^2 = 0$ contre $\mathcal{H}_1$: "$\sigma_\eta^2 > 0$. Sous $\mathcal{H}_0$, $S_n^\eta = 0$ p.s. pour tout n et le processus est stationnaire autour de sa tendance. Au contraire sous $\mathcal{H}_1$, la marche aléatoire rend le processus non-stationnaire. Le test de Kwiatkowski, Phillips, Schmidt et Shin (KPSS) présenté dans Kwiatkowski *et al.* (1992) repose sur les mêmes bases mais la partie autorégressive est remplacée par des hypothèses de mélange sur le bruit $(\varepsilon_n)_{n\geqslant 1}$ et l'existence d'une variance de long terme. Ces deux procédures possèdent un défaut crucial : elles ne prennent pas en compte les racines unitaires négatives. On se propose dans Proïa (2018) de revisiter le test LMC en enrichissant le modèle (11) au niveau de la tendance et surtout au niveau des racines unitaires, en introduisant la marche potentiellement alternée

$$\forall\, n \geqslant 1, \quad S_n^\eta = \sum_{t=1}^{n} \rho^{n-t}\, \eta_t \tag{12}$$

avec $\rho = \pm 1$. On montre qu'il existe des processus limites, construits sur le processus de Wiener, qui décrivent le comportement asymptotique de la statistique LMC correctement renormalisée, selon que l'on se place sous $\mathcal{H}_0$: "$\sigma_\eta^2 = 0$", sous $\mathcal{H}_1^+$: "$\sigma_\eta^2 > 0$ et $\rho = 1$" ou sous $\mathcal{H}_1^-$: "$\sigma_\eta^2 > 0$ et $\rho = -1$". Cela nous donne accès à un test de stationnarité dans un AR(p) avec tendance qui tient compte des racines unitaires ± 1, et qui vient donc généraliser et corriger le test LMC. Les principes d'invariance de type Donsker forment l'outil essentiel conduisant aux résultats résumés dans cette section.

Dans la continuité de la thèse

Des travaux ont été réalisés dans la continuité de la thèse mais ne s'insérant pas dans les trois axes retenus pour ce mémoire, c'est pourquoi on va se contenter ici de présenter succinctement deux d'entre eux.

Corrélation résiduelle d'ordre quelconque

La statistique de Durbin-Watson (9) n'est adaptée qu'au test de corrélation résiduelle du premier ordre : on a de l'information sur la covariance entre deux résidus consécutifs mais rien d'autre. Or dans les autorégressions d'ordre quelconque, il ne paraît pas pertinent de ramener la notion de bruit blanc à celle d'absence de corrélation du premier ordre. Dans Proïa (2018), on propose à cet égard de considérer le modèle

$$\forall\, n \geqslant 1, \quad \left\{ \begin{array}{rcl} X_n & = & \theta^T \Phi_{n-1} + \varepsilon_n \\ \varepsilon_n & = & \rho^T \Psi_{n-1} + V_n \end{array} \right. \tag{13}$$

où $\Phi_n = (X_n, \ldots, X_{n-p+1})^T$ et $\Psi_n = (\varepsilon_n, \ldots, \varepsilon_{n-q+1})^T$ pour tout $n \geqslant 0$, et où la perturbation $(V_n)_{n\geqslant 1}$ est un bruit blanc fort de moment d'ordre 4 fini. Les polynômes autorégressifs associés à θ et ρ sont supposés causaux, garantissant ainsi la stabilité du processus. Dans ce contexte, on établit le comportement asymptotique de l'estimateur des moindres carrés

de θ, qui n'est évidemment pas consistant mais reste asymptotiquement normal et de vitesse de convergence p.s. en $(\ln\ln n/n)^{1/2}$ classiquement rencontrée dans l'estimation des modèles linéaires stables. L'estimateur proposé pour ρ est donné par

$$\widehat{\rho}_n = \left(\sum_{t=1}^{n} \widehat{\Psi}_{t-1}\widehat{\Psi}_{t-1}^T\right)^{-1} \sum_{t=1}^{n} \widehat{\Psi}_{t-1}^q\, \widehat{\varepsilon}_t \tag{14}$$

avec $\widehat{\Psi}_n = (\widehat{\varepsilon}_n, \ldots, \widehat{\varepsilon}_{n-q+1})^T$ pour tout $n \geqslant 0$, où les résidus sont issus de l'estimation préalable de θ. De la même manière, on pourra ajouter une pénalisation de type *ridge* à la matrice de covariance, pour s'assurer l'inversibilité sans hypothèse supplémentaire. Le résultat principal de ce travail est que lorsque $\rho = 0$,

$$\sqrt{n}\,\widehat{\rho}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\rho^0) \tag{15}$$

avec une covariance asymptotique Σ_ρ^0 bien identifiée. Cela permet à travers une estimation consistante de Σ_ρ^0 de construire un test de $\mathcal{H}_0$: "$\rho = 0$" contre $\mathcal{H}_1$: "$\rho \neq 0$". Tenant compte de l'aspect autorégressif, la procédure obtenue est sans surprise bien plus puissante que les tests communément appliqués afin de vérifier la blancheur des résidus d'une autorégression à l'ordre quelconque. En outre, elle peut être utilisée afin de déterminer l'ordre du modèle AR qui serait donc le plus petit ne conduisant pas au rejet de $\mathcal{H}_0$. Tout comme les travaux sur la statistique de Durbin-Watson, la théorie asymptotique des martingales est intensivement utilisée pour établir ces résultats.

Distribution des résidus

Les travaux fondateurs de Bickel et Rosenblatt (1973) ont permis d'établir le comportement asymptotique de la statistique

$$\widehat{T}_n = nh_n \int_{\mathbb{R}} \big(\widehat{f}_n(x) - f(x)\big)^2 a(x)\,\mathrm{d}x \tag{16}$$

où $\widehat{f}_n$ est l'estimateur de Parzen-Rosenblatt de la densité f d'un n-échantillon, $(h_n)_{n\geqslant 1}$ est une fenêtre, selon la terminologie usuelle de la statistique non-paramétrique, et a est une fonction possédant certaines hypothèses de régularité. Il est ainsi possible d'en déduire un test d'adéquation sur la distribution de la série, au sens de la distance dans L^2. Par la suite, certains auteurs ont montré que le résultat restait vrai sous des hypothèses moins restrictives (bruits blancs, processus faiblement dépendants, etc.). Lee et Na (2002) et plus tard Bachmann et Dette (2005) valident également le résultat lorsqu'il porte sur les résidus d'un modèle autorégressif du premier ordre, à condition que ce dernier soit stable ou explosif. Spécifiquement, nous identifions dans Lagnoux *et al.* (2019), en collaboration avec A. Lagnoux et T.M.N. Nguyen, une moyenne μ et une variance asymptotique τ^2 telles que, pour les résidus d'un autorégressif d'ordre $p \geqslant 1$ stable ou explosif,

$$\frac{\widehat{T}_n - \mu}{\sqrt{h_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2) \tag{17}$$

sous des hypothèses techniques portant sur la densité f du bruit du processus, sur le noyau $\mathbb{K}$ utilisé dans l'estimation non-paramétrique de f et sur la régularité de la fonction a.

Nous montrons également que dans le cas instable pour $p = 1$, le résultat reste vrai pour la racine unitaire négative alors que, pour la racine unitaire positive, la vitesse de convergence de la statistique et son comportement asymptotique sont modifiés. Pour un $\delta > 0$ et une densité de référence f_0 qui ne s'annule pas sur $I_\delta = [-\delta, \delta]$, posons

$$\Delta_\delta(f, f_0) = \int_{I_\delta} \frac{\big(f(x) - f_0(x)\big)^2}{f_0(x)} \, \mathrm{d}x. \tag{18}$$

Nous proposons alors un protocole statistique pour tester $\mathcal{H}_0$: "$\Delta_\delta(f, f_0) = 0$" contre $\mathcal{H}_1$: "$\Delta_\delta(f, f_0) > 0$" qui met en concurrence l'hypothèse que f_0 coïncide avec f presque partout sur I_δ contre l'alternative qu'il existe un intervalle non-vide de I_δ sur lequel f_0 et f diffèrent. Ce dernier est valable pour les résidus d'autorégressions d'ordre $p \geqslant 1$ stables et explosives (et même sous quelques configurations instables), généralisant ainsi et corrigeant à certains égards les travaux de Lee et Na (2002).

Nouveaux axes de recherche

Les travaux présentés dans les sections précédentes ont un fil directeur commun : le diagnostic statistique des modèles autorégressifs (stationnarité, ordre p, blancheur et distribution résiduelles, etc.). Depuis quelques années, j'ai développé mes activités de recherche selon trois nouveaux axes quasiment orthogonaux, qui constitueront les trois chapitres de ce mémoire. Alors que les deux premiers chapitres sont essentiellement théoriques, le troisième est très appliqué voire uniquement méthodologique.

Axe 1 : autorégressifs à coefficients variables

Tout d'abord, dans la thématique autorégressive, la présence de racines unitaires a longtemps été un enjeu majeur en économétrie des séries chronologiques et en statistique des processus. En effet, comme on l'a rappelé dans la section introductive aux travaux de thèse, le processus se comporte de façon radicalement différente selon la localisation des valeurs propres de sa matrice compagnon C_θ. On sait par exemple que sous les hypothèses adéquates, si le processus est stable, alors il est asymptotiquement stationnaire et l'estimateur des moindres carrés converge à vitesse $n^{-1/2}$ vers une loi normale. Au contraire lorsqu'il est instable (avec une seule racine unitaire), il évolue comme une variable aléatoire d'ordre de grandeur $n^{1/2}$ et l'estimateur des moindres carrés converge à vitesse n^{-1} vers une loi non-gaussienne et non-symétrique, et tout cela se fait à vitesse exponentielle lorsqu'il est explosif. En résumé, il y a deux discontinuités très nettes dans le comportement du processus (en termes de valeur de X_n ou d'estimation de θ) lorsque $\rho(C_\theta)$ varie continument dans $[1 \pm \epsilon]$, pour $\epsilon > 0$. Ce changement brutal de comportement a certes motivé les procédures de recherche de racines unitaires, mais il a aussi ouvert la voie à une approche moins rigide dans laquelle les coefficients ne sont plus fixes. Un modèle autorégressif à coefficients variables peut s'écrire

$$\forall\, n \geqslant 1, \quad X_n = \theta_{n,1}\, X_{n-1} + \ldots + \theta_{n,p}\, X_{n-p} + \varepsilon_n \tag{19}$$

où $(\varepsilon_n)_{n \geqslant 1}$ est un bruit blanc et $(\theta_n)_{n \geqslant 1}$ est une suite de coefficients sur laquelle on fera des hypothèses spécifiques. Dans une telle écriture le coefficient prend une valeur

différente à chaque instant, mais on peut aussi envisager la forme triangulaire

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad X_{n,\,k} = \theta_{n,\,1}\, X_{n,\,k-1} + \ldots + \theta_{n,\,p}\, X_{n,\,k-p} + \varepsilon_k \tag{20}$$

dans laquelle les coefficients sont fixes pour un n donné, mais le modèle AR(p) est revu dans son intégralité à chaque nouvelle valeur de n. À titre d'exemple, lorsque $p = 1$ et en supposant que $|\theta_n| < 1$ mais que $|\theta_n| \to 1$ quand $n \to +\infty$, on 'zoome' en quelque sorte sur la frontière intérieure du disque unité et l'on étudie de plus près la discontinuité constatée entre $|\theta| < 1$ et $|\theta| = 1$. Phillips et Magdalinos (2007) montrent dans ce contexte que dans un voisinage d'ordre $O(n^{-\alpha})$ de la racine unitaire avec $0 < \alpha < 1$, l'estimateur converge à vitesse $n^{(-1-\alpha)/2}$ vers une loi normale. Même s'il reste des discontinuités dans la distribution asymptotique sur les bords $\alpha \to 0^+$ (où la variance asymptotique est surestimée) et $\alpha \to 1^-$ (où la forme de la distribution limite est différente), un pont est ainsi créé dans les vitesses de convergence entre la stabilité ($\alpha = 0$) et l'instabilité ($\alpha = 1$). Dans un esprit de généralisation, la première partie du Chapitre 1 sera dédiée au modèle (20) dans lequel nous établirons un principe de déviations modérées pour l'estimateur des moindres carrés sous l'hypothèse que la matrice compagnon C_{θ_n} est telle que $\rho(C_{\theta_n}) < 1$ mais que $\rho(C_{\theta_n}) \to 1$, classe de processus que nous avons qualifiés de *quasi-instables*, prolongeant en ce sens les travaux de Miao *et al.* (2015) et permettant à terme (c'est ce qui est espéré dans le cadre de nos travaux actuels) de remonter jusqu'au théorème central limite et ainsi d'étendre les résultats de Phillips et Magdalinos (2007) au cadre vectoriel. En parallèle, nous nous sommes intéressés en collaboration avec M. Soltane au cas où les coefficients sont aléatoires, configuration du modèle largement mise en lumière par Nicholls et Quinn (1981b,a), et nous avons choisi de remettre en question l'hypothèse d'indépendance temporelle entre les coefficients, qui paraît en effet peu réaliste dans un cadre chronologique. Dans la seconde partie du Chapitre 1, nous développerons un modèle similaire à (19) dans lequel les coefficients sont aléatoires et autocorrélés. Nous montrerons en particulier que l'estimation par moindres carrés n'est plus consistante et nous en donnerons le comportement asymptotique précis. Nous en tirerons également un test de corrélation dans les coefficients aléatoires, lorsque $p = 1$.

Axe 2 : modèles graphiques partiels

Par ailleurs, nous avons travaillé avec E. Okome Obiang sur une problématique de grande dimension et plus spécifiquement sur les modèles graphiques partiels gaussiens (PGGM), ramification des modèles graphiques que nous développerons en détail le moment venu. Considérons un modèle linéaire à réponses multivariées de la forme

$$\mathbb{Y} = \mathbb{X}B + E \tag{21}$$

où $\mathbb{Y} \in \mathbb{R}^{n\times q}$ contient les réponses, $\mathbb{X} \in \mathbb{R}^{n\times p}$ contient les prédicteurs, $B \in \mathbb{R}^{p\times q}$ est une matrice de coefficients et $E \in \mathbb{R}^{n\times q}$ est un bruit multivarié. L'estimation de B est un problème bien connu et largement étudié, que l'on se place en petite dimension vis-à-vis des prédicteurs ou en grande dimension ($p \gg n$) avec éventuellement structures de groupes, voir par exemple (Giraud, 2014, Chap. 6). Pour le k-ème individu, on a donc $Y_k = B^T X_k + E_k$ et dans le cas gaussien, on supposera de plus que $E_k \sim \mathcal{N}_q(0, R)$. Lorsque le vecteur des observations $(Y_k, X_k) \in \mathbb{R}^{q+p}$ est lui-même supposé suivre une distribution

$\mathcal{N}_{q+p}(0, \Sigma)$ dont la matrice de précision $\Omega = \Sigma^{-1}$ est définie positive et se décompose en blocs

$$\Omega = \begin{pmatrix} \Omega_y & \Delta \\ \Delta^T & \Omega_x \end{pmatrix} \tag{22}$$

où $\Omega_y \in \mathbb{S}^q_{++}$, $\Delta \in \mathbb{R}^{q \times p}$ et $\Omega_x \in \mathbb{S}^p_{++}$, on déduit des propriétés des vecteurs gaussiens que $Y_k \,|\, X_k \sim \mathcal{N}_q(-\Omega_y^{-1}\Delta X_k,\, \Omega_y^{-1})$. On voit ainsi se dessiner la reparamétrisation

$$B = -\Delta^T \Omega_y^{-1} \quad \text{et} \quad R = \Omega_y^{-1} \tag{23}$$

dans le modèle (21). Cependant, l'estimation de Δ est particulièrement intéressante du point de vue de l'interprétation car on sait grâce à la théorie des modèles graphiques qu'à un facteur multiplicatif près, l'élément (i, j) de Δ contient la corrélation partielle entre la i-ème réponse et le j-ème prédicteur, information que l'on ne peut généralement pas tirer directement de B. En grande dimension, l'estimation des matrices de précision gaussiennes est une problématique abondamment traitée dans la littérature récente, citons simplement dans cette introduction le célèbre *Graphical Lasso* de Friedman *et al.* (2008). Mais extraire Δ de Ω engendre une difficulté majeure lorsque p est grand, car la forte pénalisation nécessaire à l'estimation de la matrice induit un biais conséquent : on estime trop de paramètres, $(q + p)(q + p + 1)/2 = O(p^2)$, par rapport à ce dont on a réellement besoin, $q(q + p) = O(p)$. Dans un modèle graphique partiel, on va chercher à n'estimer que ce qui est utile à B, à savoir le couple (Ω_y, Δ). Dans ce contexte, une approche par vraisemblance pénalisée est proposée par Yuan et Zhang (2014) munie d'une garantie théorique, et enrichie à certains égards par Chiquet *et al.* (2017) qui y adjoignent une pénalisation structurante susceptible d'imposer des motifs dans Δ. À titre d'exemple, cela peut être particulièrement utile lorsque les prédicteurs possèdent une structure de dépendance temporelle et que l'on émet le souhait que la sélection de variables aboutisse à des segments de prédicteurs plutôt qu'à des prédicteurs isolés. Dans la première partie du Chapitre 2, nous mettrons en commun les deux études précitées afin de réfléchir à un algorithme d'estimation de (Ω_y, Δ) par maximum de vraisemblance pénalisée dans un modèle structurant, accompagné d'une garantie théorique similaire à celle de Yuan et Zhang (2014). Nous verrons en particulier que la présence d'une pénalisation structurante ne change pas l'ordre de grandeur de l'erreur d'estimation mais restreint le domaine de validité des paramètres de régularisation pour que cet ordre de grandeur soit respecté. La seconde partie du Chapitre 2 sera dédiée à la contrepartie bayésienne du modèle graphique partiel, qui à notre connaissance n'avait jamais encore été développée. Nous nous appuierons sur l'approche de Liquet *et al.* (2017), traitant l'estimation bayésienne de B dans le modèle (21), pour proposer une estimation bayésienne de (Ω_y, Δ) avec différents types de sparsité (en coordonnées et/ou en groupes) dans Δ, par une stratégie *spike-and-slab*. Nous présenterons quelques résultats issus des échantillonneurs de Gibbs qui ont été mis en place afin d'estimer la densité jointe qui découle des modèles hiérarchiques et d'obtenir des estimations *a posteriori* potentiellement génératrices de sparsité dans Δ.

Axe 3 : applications aux sciences du vivant

Enfin, nous présenterons pour conclure deux études menées afin de répondre à des besoins en analyse des données et modélisation de l'Institut de Recherche en Horticulture et

Semences (IRHS), unité mixte de recherche INRAE basée à Angers. Toutes deux traitent de populations de rosiers : la première se focalise sur un aspect phénotypique (courbes de floraison) et la seconde sur un aspect génotypique (reconstruction de généalogies). Dans la première partie du Chapitre 3, nous mettons en place des mélanges gaussiens afin de modéliser les courbes de floraison des rosiers sur l'année (mesurées en densité de fleurs sur la plante au cours du temps). Nous en tirons des indicateurs permettant un clustering de la population et des profils caractéristiques de floraison dans chaque groupe. Contrairement à la première, la seconde étude n'est pas une analyse statistique mais une construction probabiliste de graphes. À partir de marqueurs génétiques sur les individus et d'autres informations de nature descriptive (comme la date d'obtention de la plante ou sa ploïdie), on cherche à reconstruire rétrospectivement un arbre généalogique de la population sur le modèle de Chaumont *et al.* (2017) mais dans un contexte plus général. En effet, le monde végétal est à l'origine de nombreuses difficultés dans cette étude : les individus de la population sont di-, tri- ou tétraploïdes (les chromosomes vont par 2, 3, ou 4), ce qui engendre des schémas de reproduction bien plus compliqués que le schéma standard $\{a, b\} \times \{c, d\} \mapsto \{ac, ad, bc, bd\}$ muni de la probabilité uniforme. Par ailleurs, certaines données génotypiques ne sont pas connues avec certitude en présence de tri- ou tétraploïdie. Tenant compte de cela, on proposera dans la seconde partie du Chapitre 3 un travail en collaboration avec C. Trabelsi axé sur un algorithme de reconstruction généalogique permettant de mettre en évidence l'arbre de maximum de vraisemblance, une estimation de la loi de reproduction des individus de la population et donc des individus atypiques (probablement favorisés par les sélectionneurs) ou encore une tentative d'identification des chaînons manquants (nœuds de l'arbre disparus de l'étude).

Bibliographie

Bachmann, D. et Dette, H. (2005). A note on the the Bickel-Rosenblatt test in autoregressive time series. *Stat. Probab. Lett.*, 74:221–234.

Bercu, B. et Proïa, F. (2013). A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. *ESAIM Probab. Stat.*, 17:500–530.

Bercu, B., Proïa, F. et Savy, N. (2014). On Ornstein-Uhlenbeck driven by Ornstein-Uhlenbeck processes. *Stat. Probab. Lett.*, 85:36–44.

Bickel, P. J. et Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1:1071–1095.

Bitseki Penda, V., Djellout, H. et Proïa, F. (2014). Moderate deviations for the durbin-watson statistic related to the first-order autoregressive process. *ESAIM Probab. Stat.*, 18: 308–331.

Brockwell, P. J. et Davis, R. A. (2006). *Time series : theory and methods*. Springer Series in Statistics. Springer, New York.

Chan, N. H. et Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.*, 16:367–401.

Chaumont, L., Malécot, V., Pymar, R. et Sbai, C. (2017). Reconstructing pedigrees using probabilistic analysis of ISSR amplification. *J. Theor. Biol.*, 412:8–16.

Chiquet, J., Mary-Huard, T. et Robin, S. (2017). Structured regularization for conditional Gaussian graphical models. *Stat. Comput.*, 27(3):789–804.

Duflo, M. (1997). *Random iterative models*. Applications of Mathematics (vol. 34), New York. Springer-Verlag, Berlin.

Durbin, J. et Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, 37:409–428.

Durbin, J. et Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, 38:159–178.

Durbin, J. et Watson, G. S. (1971). Testing for serial correlation in least squares regession. III. *Biometrika*, 58:1–19.

Friedman, J., Hastie, T. et Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics.*, 9(3):432–441.

Giraud, C. (2014). *Introduction to High-Dimensional Statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. et Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root ? *J. Econometrics.*, 54:159–178.

Lagnoux, A., Nguyen, T. M. N. et Proïa, F. (2019). On the Bickel-Rosenblatt test of goodness-of-fit for the residuals of autoregressive processes. *ESAIM Probab. Stat.*, 23:464–491.

Lai, T. L. et Wei, C. Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.*, 13:1–23.

Lee, S. et Na, S. (2002). On the Bickel-Rosenblatt test for first-order autoregressive models. *Stat. Probab. Lett.*, 56:23–35.

Leybourne, S. J. et McCabe, B. P. M. (1994). A consistent test for a unit root. *J. Bus. Econ. Stat.*, 12:157–166.

Liquet, B., Mengersen, K., Pettitt, A. N. et Sutton, M. (2017). Bayesian variable selection regression of multivariate responses for group data. *Bayesian Anal.*, 12(4):1039–1067.

Miao, Y., Wang, Y. et Yang, G. (2015). Moderate deviation principles for empirical covariance in the neighbourhood of the unit root. *Scand. J. Stat.*, 42:234–255.

Nicholls, D. F. et Quinn, B. G. (1981a). The estimation of multivariate random coefficient autoregressive models. *J. Multivar. Anal.*, 11:544–555.

Nicholls, D. F. et Quinn, B. G. (1981b). Multiple autoregressive models with random coefficients. *J. Multivar. Anal.*, 11:185–198.

Phillips, P. C. B. et Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *J. Econometrics.*, 136:115–130.

Proïa, F. (2013). Further results on the H-Test of Durbin for stable autoregressive processes. *J. Multivariate Anal.*, 118:77–101.

Proïa, F. (2018). Stationarity against integration in the autoregressive process with polynomial trend. *Probab. Math. Stat.*, 38:1–26.

Proïa, F. (2018). Testing for residual correlation of any order in the autoregressive process. *Commun. Stat. A-Theor.*, 47:628–654.

Samoura, Y. (2017). Moderate deviations for the Durbin-Watson statistic associated to the stable p-order autoregressive process. *Far East J. Theor. Stat.*, 53:349–389.

Yuan, X. T. et Zhang, T. (2014). Partial Gaussian graphical model estimation. *IEEE. T. Inform. Theory.*, 60(3):1673–1687.

Chapitre 1

Autorégressifs à coefficients variables

Comme nous l'avons mentionné dans l'introduction, ce chapitre est dévolu aux processus autorégressifs à coefficients variables. Nous traiterons tout d'abord des processus quasi-instables, avant d'évoquer dans un second temps les processus à coefficients aléatoires. Nous donnons le contenu explicite de deux articles, un par section, et nous renvoyons à l'introduction (en particulier le résumé des travaux de thèse) pour une présentation générale du contexte autorégressif (estimation, stabilité, etc.).

1.1 Processus quasi-instables

Nous présentons dans cette section le contenu de l'article Proïa (2020), publié dans *Journal of Statistical Planning and Inference* et dont l'objectif est l'établissement d'un principe de déviations modérées (PDM) sur l'erreur d'estimation dans une classe de processus autorégressifs quasi-instables. Nous voyons cela comme un premier pas avant de remonter jusqu'au théorème central limite.

Résumé

Soit le processus autorégressif d'ordre $p \geqslant 1$ à coefficients variables donné par

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \Phi_{n,\,k} = A_n\, \Phi_{n,\,k-1} + E_k \tag{1.1}$$

où $E_k = (\varepsilon_k, 0, \ldots, 0)^T$ est un bruit blanc fort p-vectoriel, où $\Phi_{n,\,k} = (X_{n,\,k}, \ldots, X_{n,\,k-p+1})^T$ avec comme valeur initiale $\Phi_{n,0}$, et avec A_n comme matrice compagnon (telle que définie dans l'introduction). Pour une série chronologique de taille $n \geqslant 1$, l'estimateur des moindres carrés de θ_n est donné par

$$\widehat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,\,k-1}\, X_{n,\,k} \quad \text{avec} \quad S_{n-1} = \sum_{k=1}^{n} \Phi_{n,\,k-1}\, \Phi_{n,\,k-1}^T \tag{1.2}$$

quitte à ajouter un petit élément diagonal pour s'assurer de l'inversibilité de S_{n-1}. La principale hypothèse émise sur les coefficients est que $\forall\, n \geqslant 1$, $\rho(A_n) < 1$ mais que $\rho(A_n) \to 1$. Ainsi le processus est stable mais asymptotiquement instable, le rayon spectral de sa matrice compagnon se situe dans un voisinage de la frontière intérieure du disque unité. Lorsque $A_n = A$ dans le cas stable où $\rho(A) < 1$, (Worms, 1999, Thm. 3) a montré

que la suite formée par l'erreur d'estimation correctement renormalisée satisfaisait un principe de grandes déviations (PGD). Le résultat principal de notre étude est qu'il en va de même dans le cas quasi-instable sous quelques hypothèses techniques, en particulier le fait que la convergence de $\rho(A_n)$ vers 1 ne doit se faire au mieux qu'à vitesse polynomiale. Explicitement, on construit une classe de vitesses $(b_n)_{n\geqslant 1}$ et une fonction $I_\theta : \mathbb{R}^p \to \bar{\mathbb{R}}^+$ telles que

$$\left(\frac{\sqrt{n}}{b_n \sqrt{1-\rho(A_n)}} \left(\widehat{\theta}_n - \theta_n \right) \right)_{n\geqslant 1} \tag{1.3}$$

satisfait un PGD de vitesse $(b_n^2)_{n\geqslant 1}$ et de taux I_θ, à condition que la matrice de covariance asymptotique correctement renormalisée soit inversible. Dans le cas contraire, on montre un résultat similaire pour une estimation pénalisée (de type *ridge*). On prolonge ainsi les travaux de Miao *et al.* (2015), dédiés au cas $p = 1$. Une analyse fine de la matrice A_n est faite, en particulier on construit sa diagonalisation (pour n suffisamment grand) à l'aide d'une matrice de passage de Vandermonde et son inverse, dont les coefficients sont reliés aux polynômes d'interpolation de Lagrange. Cela permet en particulier de montrer que ces matrices sont bornées (pour une quelconque norme), argument décisif dans le fait que la croissance de $\|A_n\|$ se rattache à celle de $\rho(A_n)$ et que l'on puisse faire apparaître directement $\rho(A_n)$ dans la vitesse des déviations modérées. Par la suite, les techniques de preuves utilisées reposent essentiellement sur les séries (m_n)-dépendantes (avec $m_n \to +\infty$) et sur le théorème de Gärtner-Ellis (Dembo et Zeitouni, 1998, Sec. 2.3). L'article est fourni en fin de section.

Perspectives

Cette étude couvre l'essentiel de la problématique du PDM pour l'erreur d'estimation dans un modèle quasi-instable (au sens que nous avons donné à cet adjectif), mais n'en couvre pas non plus l'intégralité. En particulier, nos démonstrations font apparaître des difficultés lorsqu'il existe des racines unitaires de multiplicité supérieure à deux dans le processus asymptotique ou, de manière équivalente, lorsqu'il existe au moins deux valeurs propres de A, limite de A_n, égales et de module 1, et tout laisse à penser que la vitesse des déviations doit évoluer dans ce cas. En soi c'est un comportement attendu car, comme on l'a rappelé en introduction, on sait depuis Chan et Wei (1988) que le nombre de racines unitaires joue sur la vitesse de convergence de l'estimation. Cette extension devrait être possible sans grandes complications, par contre établir un PGD pour l'erreur d'estimation se révèlerait un problème bien plus compliqué, d'autant qu'à notre connaissance il n'est pas entièrement établi dans le cas stable où $\rho(A_n) = \rho(A) < 1$, ce qui révèle clairement la difficulté d'une telle l'étude. Peut-être de manière plus naturelle, nos résultats permettent d'établir la consistance faible de l'estimation, c'est-à-dire que

$$\left\| \widehat{\theta}_n - \theta_n \right\| \xrightarrow{\mathbb{P}} 0 \tag{1.4}$$

mais pas la consistance forte. Les techniques de preuve de Lai et Wei (1983) qui montrent la convergence p.s. dans les cas stable et instable devraient nous permettre de l'établir dans notre contexte intermédiaire, mais la piste reste à creuser. Enfin, on a vu que lorsque $p = 1$, Phillips et Magdalinos (2007) montrent que dans un voisinage d'ordre $O(n^{-\alpha})$ de la racine unitaire avec $0 < \alpha < 1$, l'estimateur converge à vitesse $n^{(-1-\alpha)/2}$ vers une

loi normale. Par analogie, on peut conjecturer que la suite (1.3) est asymptotiquement normale lorsque $b_n = 1$, c'est un travail qui justement est en cours (en collaboration avec M. Badreau). En particulier, il semble qu'il existe une variance Σ telle que

$$\frac{\sqrt{n}}{\sqrt{1-\rho(A_n)}}\left(\widehat{\theta}_n - \theta_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p(0,\Sigma) \tag{1.5}$$

et nous cherchons à la caractériser par des techniques de martingales. Buchmann et Chan (2013) proposent une théorie unifiée dans un contexte plus large, cependant la quasi-instabilité y est introduite à travers une perturbation de la jordanisation de la matrice compagnon, et donc pas de la même manière que dans notre travail. C'est un fil qu'il pourrait être intéressant de suivre également, car on trouve dans cette référence les distributions asymptotiques de l'estimation sous ces hypothèses peu faciles à manipuler.

MODERATE DEVIATIONS IN A CLASS OF STABLE BUT NEARLY UNSTABLE PROCESSES

FRÉDÉRIC PROÏA

Abstract. We consider a stable but nearly unstable autoregressive process of any order. The bridge between stability and instability is expressed by a time-varying companion matrix A_n with spectral radius $\rho(A_n) < 1$ satisfying $\rho(A_n) \to 1$. In that framework, we establish a moderate deviation principle for the empirical covariance only relying on the elements of A_n through $1 - \rho(A_n)$ and, as a by-product, we establish a moderate deviation principle for the OLS estimator when Γ, the renormalized asymptotic variance of the process, is invertible. Finally, when Γ is singular, we also provide a compromise in the form of a moderate deviation principle for a penalized version of the estimator. Our proofs essentially rely on truncations and deviations of m_n–dependent sequences, with an unbounded rate (m_n).

1. Introduction and Assumptions

Unit root issues have long been crucial in time series econometrics and have therefore focused a great deal of research studies. This sudden demarcation between stability and instability is responsible for many inference problems in linear time series (see Brockwell and Davis [4] for a detailed overview of the linear stochastic processes). The remarkable works of Chan and Wei [7] encompass, in a much more general context, the now well-known fact that the least squares estimator is $\sqrt{n}$–consistent with Gaussian behavior when the underlying autoregressive process is stable, whereas it is n–consistent with asymmetrical distribution when the process is unstable. This rather abrupt change in the rate of convergence and in the asymptotic distribution certainly motivated the wide range of unit root testing procedures, but it also paved the way for studies based on time-varying coefficients. In a nearly unstable autoregressive process, we do not focus on a parameter θ satisfying $|\theta| < 1$ or $|\theta| = 1$ but, instead, the parameter is considered as a sequence (θ_n) such that $|\theta_n| < 1$ and $|\theta_n| \to 1$ as $n \to +\infty$. This sample size dependent structure allows a continuity between stability and instability. For example, Phillips and Magdalinos [20] treat the case where the coefficient is in a $O(\kappa_n^{-1})$ neighborhood of the unit root with $\kappa_n = n^\alpha = o(n)$. Amongst other results, they prove a central limit theorem for the estimator at the rate $\sqrt{n\,\kappa_n}$, thereby making a bridge between the stable rate $\sqrt{n}$ and the unstable rate n. In the same vein, let us also mention the work of Chan and Wei [6], natural generalizations like the study of Phillips and Lee [19] related to vector autoregressions, or the recent unified theory of Buchmann and Chan [5], focused on nearly unstable autoregressive processes. Our paper is precisely based on the latter topic, in a sense that will be precised in good time.

Given a parametric generating process, the precision of the estimation is usually assessed by its rate of convergence and the deviations can be seen as a natural continuation after a central limit theorem or even a law of iterated logarithm. Roughly speaking, they may

Key words and phrases. Nearly unstable autoregressive process, Moderate deviation principle, OLS estimation, Asymptotic behavior, Unit root.

be used to estimate the exponential decline of the probability of tail events related to the distance between the estimator and the parameter of interest. We refer to Dembo and Zeitouni [8] regarding the mathematical formalization. Since the 1980s, numerous authors have worked on large and/or moderate deviations in a time series context under many and varied hypotheses. Without claiming to be exhaustive, one can mention the studies of Donsker and Varadhan [10] and Bercu *et al.* [2] on stationary Gaussian processes and quadratic forms, the paper of Worms [21] on Markov chains and regression models and the one of Bercu [1] on first-order Gaussian stable, unstable and explosive processes. One can also mention the works of Mas and Menneteau [15] on Hilbertian processes, Djellout *et al.* [9] on non-linear functionals of moving average processes, Wu and Zhao [22] on stationary non-linear processes, Miao and Shen [16] on general autoregressive processes or, more recently, Bitseki Penda *et al.* [3] on first-order processes with correlated errors. All the references inside may complete this concise list.

In this paper, we investigate the moderate deviations of the estimate in stable but nearly unstable autoregressions. This can be seen as a full generalization of the recent work of Miao, Wang and Yang [17], focused on the univariate case. Our proofs essentially rely on truncations and deviations of m_n–dependent sequences where the rate (m_n) is unbounded. The main technical contributions are twofold. On the one hand, expressing the nearly instability directly through the sequence of spectral radii of the companion matrix seems, to the best of our knowledge, a new approach having many advantages. For example the authors of the recent paper [5] introduce a perturbation in the Jordan canonical form of the model (see Thm. 2.1) which is a powerful idea to deal with the subject of their study, but somehow unnecessarily complex for ours. On the other hand, from a purely technical point of view, unbounded truncations have already been used to get moderate deviations (see *e.g.* [18] and [17]), but we will see that the vector case treated here and the specific features of the model cannot be adapted as easily to the existing tools. As a consequence, we need to redevelop a full Gärtner-Ellis reasoning to establish the deviations of our unbounded vector truncations. This quite general strategy might inspire future similar studies.

For a fixed $n \geqslant 1$, let the process be given for some $p \geqslant 1$ and $k \in \{1, \ldots, n\}$ by

$$X_{n,\,k} = \sum_{i=1}^{p} \theta_{n,\,i}\, X_{n,\,k-i} + \varepsilon_k$$

where $(\varepsilon_k)_k$ is a sequence of zero-mean i.i.d. random variables. In an equivalent way, we can consider the vector expression

$$\Phi_{n,\,k} = A_n\, \Phi_{n,\,k-1} + E_k \tag{1.1}$$

where $E_k = (\varepsilon_k, 0, \ldots, 0)^T$ is a p–vectorial noise, $\Phi_{n,\,k} = (X_{n,\,k}, \ldots, X_{n,\,k-p+1})^T$ and

$$A_n = \begin{pmatrix} \theta_{n,\,1} & \theta_{n,\,2} & \ldots & \theta_{n,\,p} \\ & I_{p-1} & & 0 \end{pmatrix} \tag{1.2}$$

is the $p \times p$ companion matrix of the autoregressive process. If $(E_k)_k$ has a finite variance, it is well-known that $(\Phi_{k,\,n})_k$ is a second-order stationary process having the causal form

$$\Phi_{n,\,k} = \sum_{\ell=0}^{+\infty} A_n^{\ell}\, E_{k-\ell} \tag{1.3}$$

when $\rho(A_n) < 1$, that is, when the largest modulus of its eigenvalues is less than 1 (see *e.g.* Thm. 11.3.1 of [4] and the fact that each eigenvalue of A_n is the inverse of a zero of the autoregressive polynomial of the process). Since $(\varepsilon_k)_k$ is an i.i.d. sequence, the process is strictly stationary with mean zero and variance given by

$$\Gamma_n = \sigma^2 \sum_{\ell=0}^{+\infty} A_n^\ell \, K_p \, (A_n^T)^\ell \tag{1.4}$$

where, for convenience, we will denote in the whole study

$$K_p = \begin{pmatrix} 1 & 0 \\ 0 & 0_{p-1} \end{pmatrix} \quad \text{and} \quad U_p = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{1.5}$$

the $p \times p$ matrix with 1 at the top left and 0 elsewhere, and its first column standing for the first vector of the canonical basis of $\mathbb{R}^p$. As a consequence of the causal expression above, the initial vector $\Phi_{n,\,0}$ is not arbitrary and has to share the distribution of the process. This also implies the relation

$$\Gamma_n = A_n \, \Gamma_n \, A_n^T + \sigma^2 \, K_p. \tag{1.6}$$

As will be largely developped throughout the study, Γ_n is finite for all $n \geqslant 1$ but, as n increases, $|||\Gamma_n||| \to +\infty$. The keystone matrix Γ obtained after a correct standardization of Γ_n is the renormalized asymptotic variance of the process. Before we start, we define a matrix that will also prove to be crucial to our results,

$$B_n = I_{p^2} - A_n \otimes A_n. \tag{1.7}$$

We are now going to introduce and comment the hypotheses that will be needed, though not always simultaneously, in the whole paper. Section 2 is devoted to our main results : two statements related to the moderate deviations of the empirical covariance and the OLS estimator, a set of explicit examples and some additional comments and conclusions. Finally, in Section 3 divided into numerous subsections, we will prove all our results, step by step.

Remark. We denote by $\|\cdot\|$ the Euclidean vector norm and by $|||\cdot|||$ the spectral matrix norm. Other norms may be used, in which case an appropriated subscript is added. Moreover, we will always denote by $\langle \cdot, \cdot \rangle$ the usual inner product of the Euclidean space $\mathbb{R}^d$ for any $d \geqslant 1$. We write $M^\dagger$ for the Moore-Penrose pseudo-inverse of any matrix M (see *e.g.* Sec. 0.3 of [12]).

1.1. **Hypotheses.** First of all, we present the hypotheses that we retain.

(H_1) *Gaussian integrability condition.* There exists $\alpha > 0$ such that

$$\mathbb{E}\big[\mathrm{e}^{\alpha \, \varepsilon_1^2}\big] < +\infty$$

where ε_1 represents the zero-mean i.i.d. sequence $(\varepsilon_k)_k$ of variance $\sigma^2 > 0$ and fourth-order moment $\tau^4 > 0$.

(H_2) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \to +\infty} A_n = A$$

with distinct eigenvalues $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_1| = \rho(A)$, and the top right element of A is non-zero.

(H$_3$) *Spectral radius of the companion matrix.* For all $n \geqslant 1$, $\rho(A_n) < 1$. In addition,

$$\lim_{n \to +\infty} \rho(A_n) = \rho(A) = 1.$$

(H$_4$) *Renormalization.* We have the convergences

$$\lim_{n \to +\infty} \frac{B_n^{-1}}{|||B_n^{-1}|||_*} = H \quad \text{and} \quad \lim_{n \to +\infty} (1 - \rho(A_n)) \, |||B_n^{-1}|||_* = h$$

for some matrix norm, where H is a $p^2 \times p^2$ non-zero matrix and $h > 0$.

(H$_5$) *Moderate deviations.* The moderate deviations scale (b_n) satisfies

$$\lim_{n \to +\infty} b_n = +\infty \quad \text{and} \quad \lim_{n \to +\infty} \frac{\sqrt{n} \, (1 - \rho(A_n))^{\frac{3}{2}+\eta}}{b_n} = +\infty$$

for a small $\eta > 0$.

1.2. **Comments on the hypotheses.** First, assuming in (H$_2$) that the limiting matrix has distinct eigenvalues is a matter of simplication of the reasonings. Indeed, A_n turns out to be diagonalizable for a sufficiently large n, and, as a companion matrix, it is well-known that the change of basis is done *via* a Vandermonde matrix having numerous nice properties (more details are given in Section 3.1, and a discussion on the case of multiple eigenvalues is provided in Section 2.3). The top right element of A_n is $\theta_{n,\,p}$. So, assuming in (H$_2$) that $\theta_{n,\,p} \nrightarrow 0$ ensures that the limit process is still of order p and that 0 cannot be an eigenvalue of A, since $\det(A) = (-1)^{p+1} \, \theta_p$. Moreover, note that, in (H$_4$), the invertibility of B_n for all n is guaranteed by (H$_3$). Indeed, $\rho(A_n \otimes A_n) = \rho^2(A_n) < 1$ (see *e.g.* Lem. 5.6.10 and Cor. 5.6.16 of [13]). In addition, we obviously have, for all $\ell \geqslant 0$,

$$\rho(A_n^\ell) = \rho^\ell(A_n) \leqslant |||A_n^\ell|||$$

so that we get

$$\frac{1}{1 - \rho(A_n)} \leqslant \sum_{\ell=0}^{+\infty} |||A_n^\ell||| = L_n \tag{1.8}$$

giving a lower bound for L_n. Similarly,

$$\frac{1}{(1 - \rho(A_n))^2} \leqslant \sum_{\ell=0}^{+\infty} (\ell + 1) \, |||A_n^\ell||| = M_n. \tag{1.9}$$

However, an exact upper bound for these sums may be difficult to reach and may require stringent conditions on the elements of A_n. We refer the reader to Lemma 3.1 where, under (H$_2$) and (H$_3$), some asymptotic upper bounds are established. We also refer to Section 2.2 where the explicit calculations in terms of some examples shall help to understand the rates involved in the hypotheses. Now for a fixed $n \geqslant 1$, let

$$\mu_n = \rho(A_n) + \frac{1 - \rho(A_n)}{2} = \frac{\rho(A_n) + 1}{2}.$$

Clearly, $\rho(A_n) < \mu_n < 1$. Hence, according to Prop. 2.3.15 of [11], for all $n \geqslant 0$, there exists a constant $c_n > 0$ such that, for all $\ell \geqslant 0$, $|||A_n^\ell||| \leqslant c_n \, \mu_n^\ell$ so that

$$L_n \leqslant \frac{c_n}{1 - \mu_n} < +\infty \quad \text{and} \quad M_n \leqslant \frac{c_n}{(1 - \mu_n)^2} < +\infty.$$

Letting n tend to infinity, it follows from (H_3) and (H_4) that

$$\lim_{n \to +\infty} |||B_n^{-1}||| = \lim_{n \to +\infty} L_n = \lim_{n \to +\infty} M_n = +\infty. \tag{1.10}$$

Finally, it will be established in good time that there is a limiting matrix Γ such that

$$\lim_{n \to +\infty} \frac{\Gamma_n}{|||B_n^{-1}|||_*} = \Gamma \tag{1.11}$$

where $|||\cdot|||_*$ is the matrix norm of (H_4).

Remark. To facilitate the reading, we consider from now on that the matrix norm $|||\cdot|||_*$ is identified in (H_4), and we will only note $|||\cdot|||$ in what follows.

2. Main results

This section contains two statements that constitute the main results of the paper. The first of them is quite long to establish and will need numerous technical lemmas, but the second one will essentially be deduced as a corollary of the first one. Subsequently, we provide some explicit examples for a better understanding and an easier interpretation of the hypotheses together with some graphics showing the evolution of the processes and the estimation of the autoregressive parameter. At the end of the section, we discuss the case of multiple eigenvalues. But, first, let us recall the definition of the large and moderate deviation principles (see Sec. 1.2 of [8] for more details). In what follows, a speed is considered as a positive sequence increasing to infinity.

Definition. A sequence of random variables $(U_n)_n$ on a topological space $(\mathcal{X}, \mathcal{B})$ satisfies a large deviation principle (LDP) with speed (a_n) and rate I if there is a lower semicontinuous mapping $I : \mathcal{X} \to \bar{\mathbb{R}}^+$ such that :

- for any closed set $F \in \mathcal{B}$,

$$\limsup_{n \to +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in F) \leqslant - \inf_{x \in F} I(x),$$

- for any open set $G \in \mathcal{B}$,

$$- \inf_{x \in G} I(x) \leqslant \liminf_{n \to +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in G).$$

In particular, if the infimum of I coincides on the interior H° and the closure $\bar{H}$ of some $H \in \mathcal{B}$, then

$$\lim_{n \to +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in H) = - \inf_{x \in H} I(x).$$

Definition. A sequence of random variables $(V_n)_n$ on a topological space $(\mathcal{X}, \mathcal{B})$ satisfies a moderate deviation principle (MDP) with speed (b_n^2) and rate I if there is a speed (v_n) with $\frac{v_n}{b_n} \to +\infty$ such that $(\frac{v_n}{b_n} V_n)_n$ satisfies a large deviation principle of speed (b_n^2) and rate I.

2.1. **Moderate deviations.** We now consider an observable trajectory $X_{n,-p+1},\ldots,X_{n,n}$ for some fixed $n \geqslant 1$, and use it to provide an estimation of the parameter. It is well-known that the ordinary least squares (OLS) estimator of $\theta_n = (\theta_{n,1},\ldots,\theta_{n,p})^T$ is given by

$$\widehat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,k-1}\, X_{n,k} \quad \text{where} \quad S_{n-1} = \sum_{k=1}^{n} \Phi_{n,k-1}\, \Phi_{n,k-1}^T. \tag{2.1}$$

The first result is dedicated to the empirical variance $\frac{S_n}{n}$.

Theorem 2.1. *Under hypotheses (H_1)–(H_5), the sequence*

$$\left(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n} \operatorname{vec}\left(\frac{1}{n} \sum_{k=1}^{n} (\Phi_{n,k}\, \Phi_{n,k}^T - \Gamma_n) \right) \right)_{n \geqslant 1}$$

satisfies an LDP with speed (b_n^2) and a rate function $I_\Gamma : \mathbb{R}^{p^2} \to \bar{\mathbb{R}}^+$ defined as

$$I_\Gamma(x) = \begin{cases} \frac{1}{2\,h^3} \left\langle x, \Upsilon^{\dagger}\, x \right\rangle & \text{for } x \in \operatorname{Im}(\Upsilon) \\ +\infty & \text{otherwise} \end{cases}$$

where Υ is explicitely given in (3.18) *and h comes from (H_4).*

Proof. See Section 3.2.5. □

Remark. Through vectorization, this MDP is established on $\mathbb{R}^{p^2}$ in order to avoid any confusion in the notations, but we might work in $\mathbb{R}^{p\times p}$ as well. The associated rate function would only require a slight modification of the proof.

Remark. To be punctilious, we may add a small $\epsilon > 0$ to the diagonal of S_{n-1} to ensure that it is non-sigular for all $n \geqslant 1$ without disturbing the asymptotic behavior.

When the variance Γ given in (1.11) is invertible, we establish the MDP for the OLS in the theorem that follows. However, when it is not the case, there are some technical complications and, to reach an intermediate result, we need to introduce a penalized version of the OLS. For a small $\pi \geqslant 0$, define

$$\widehat{\theta}_n^{\,\pi} = (S_{n-1}^{\pi})^{-1} \sum_{k=1}^{n} \Phi_{n,k-1}\, X_{n,k} \quad \text{where} \quad S_{n-1}^{\pi} = S_{n-1} + \pi\, n\, |||B_n^{-1}|||\, I_p \tag{2.2}$$

with possibly $\pi = 0$ if Γ is invertible, in which case it is clearly the standard OLS given above, but necessarily $\pi > 0$ otherwise. Consider also the penalized version of the variance and the corrected parameter

$$\Gamma_\pi = \Gamma + \pi\, I_p \quad \text{and} \quad \theta_n^{\pi} = (S_{n-1}^{\pi})^{-1} S_{n-1}\, \theta_n. \tag{2.3}$$

By construction, Γ is, at worst, non-negative definite and for $\pi > 0$, Γ_π turns out to be invertible. The same goes for S_{n-1}^{π}.

Corollary 2.2. *Under hypotheses (H_1)–(H_5), for all $\pi > 0$, the sequence*

$$\left(\frac{\sqrt{n}}{b_n\,(1-\rho(A_n))^{\frac{1}{2}}} \left(\widehat{\theta}_n^{\,\pi} - \theta_n^{\pi} \right) \right)_{n \geqslant 1}$$

satisfies an LDP with speed (b_n^2) *and a rate function* $I_\theta^\pi : \mathbb{R}^p \to \bar{\mathbb{R}}^+$ *defined as*

$$I_\theta^\pi(x) = \begin{cases} \frac{h}{2\,\sigma^2}\,\langle x, \Gamma_\pi\,\Gamma^\dagger\,\Gamma_\pi\,x\rangle & \text{for } x \in \mathrm{Im}(\Gamma_\pi^{-1}\,\Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

where the variance Γ *is given in* (1.11), Γ_π *is the penalized variance given in* (2.3) *and* h *comes from* (H_4)*, respectively. If in addition* Γ *is invertible, then the sequence*

$$\left(\frac{\sqrt{n}}{b_n\,(1-\rho(A_n))^{\frac{1}{2}}}\,\big(\widehat{\theta}_n - \theta_n\big)\right)_{n\geqslant 1}$$

satisfies an LDP with speed (b_n^2) *and a rate function* $I_\theta : \mathbb{R}^p \to \mathbb{R}^+$ *defined as*

$$I_\theta(x) = \frac{h}{2\,\sigma^2}\,\langle x, \Gamma\,x\rangle.$$

Proof. See Section 3.2.6. □

To sum up, this result shows that, when Γ is invertible, the OLS satisfies an MDP, and even when Γ is singular, one may reach a compromise by getting an MDP for a penalized estimation. In the same vein, notice also that, in the invertible case,

$$\lim_{\pi\,\to\,0^+} I_\theta^\pi(x) = I_\theta(x).$$

Remark. In the stable case where $\rho(A_n) = \rho(A) < 1$, we simply have $(1-\rho(A_n))\,|||B_n^{-1}||| = h$ and $\Gamma_n\,|||B_n^{-1}|||^{-1} = \Gamma$ for all $n \geqslant 1$. By contraction, the MDP of Corollary 2.2 coincides with the one of Thm. 3 of [21] when Γ is invertible.

2.2. **Some explicit examples.** Before giving some examples, we can already note that (H_5) implies $\sqrt{n}\,(1-\rho(A_n)) \to +\infty$. Thus, necessarily, the convergence $1-\rho(A_n) \to 0$ cannot occur with an exponential rate, this is the reason why we focus on polynomial rates of the form $1-\rho(A_n) = c\,n^{-\alpha}$ for some $c > 0$ in this section. Accordingly, in all the examples below, (H_5) is only possible when $0 < \alpha < \frac{1}{3+2\eta} < \frac{1}{3}$. Thus, one cannot expect a sequence of coefficients moving too fast toward instability. The domain of validify of the speed of the MDP will be

$$1 \ll b_n \ll n^{\frac{1-(3+2\eta)\alpha}{2}} \ll \sqrt{n}.$$

2.2.1. *Univariate case with one nearly unit root.* Suppose that $p = 1$. Then, (H_2) and (H_3) imply that $|\theta_n| < 1$ and $\theta_n \to \pm 1$. We also have $B_n = 1-\theta_n^2$ and (H_4) can be expressed like

$$\lim_{n\,\to\,+\infty} \frac{B_n^{-1}}{|B_n^{-1}|} = 1 \quad \text{and} \quad \lim_{n\,\to\,+\infty} (1-|\theta_n|)\,|B_n^{-1}| = \frac{1}{2}.$$

A straightforward calculation shows that

$$\Gamma_n = \frac{\sigma^2}{1-\theta_n^2} \quad \text{and} \quad \Gamma = \sigma^2 > 0$$

so that we can choose $\pi = 0$. The standard cases, illustrated on Figure 1, are $\theta_n = 1 - c_1\,n^{-\alpha}$ for the positive unit root and $\theta_n = -1 + c_2\,n^{-\alpha}$ for the negative unit root, with $c_1, c_2 > 0$ and $\alpha > 0$. The rate function associated with Corollary 2.2 is $I_\theta(x) = \frac{x^2}{4}$, which corresponds to Prop. 2.1 of [17]. Indeed, their rate $x \mapsto \frac{x^2}{2}$ is associated to an LDP with the renormalization $(1-\theta_n^2)^{\frac{1}{2}}$ whereas our normalization is $(1-|\theta_n|)^{\frac{1}{2}}$. By contraction, the asymptotic factor $\sqrt{2}$ explains the difference.

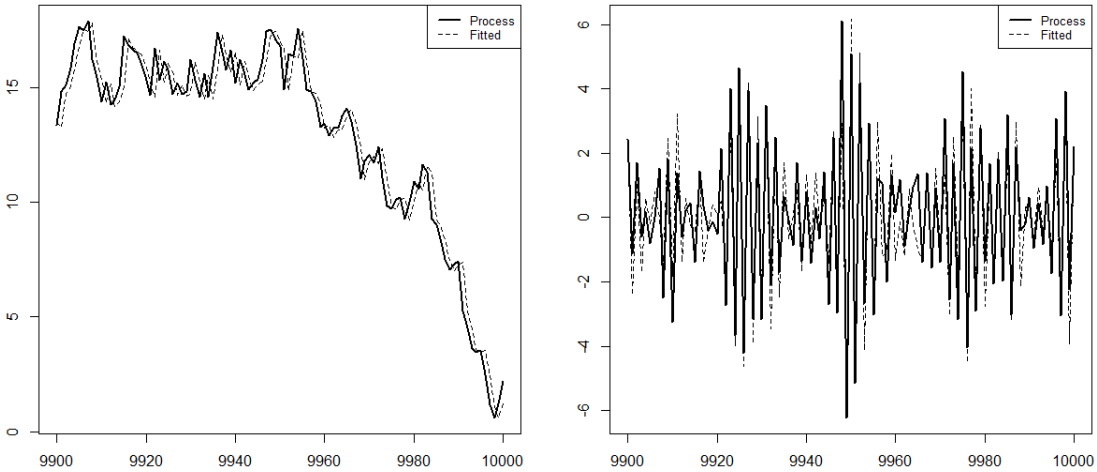


FIGURE 1. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 0$ and $\mathcal{N}(0,1)$ innovations. The setting is $c_1 = 0.1$ and $\alpha = 0.32$ on the left, $c_2 = 0.1$ and $\alpha = 0.32$ on the right.

2.2.2. *Bivariate case with one nearly unit root.* Suppose now that $p = 2$ and $\mathrm{sp}(A) = \{\pm 1, \lambda\}$ with $|\lambda| < 1$. This situation occurs, for example, when

$$A_n = \begin{pmatrix} \lambda + 1 - c\,n^{-\alpha} & -\lambda\,(1 - c\,n^{-\alpha}) \\ 1 & 0 \end{pmatrix}$$

whose eigenvalues are $1 - c\,n^{-\alpha}$ and λ. This is illustrated on Figure 2. For $c > 0$ and $\alpha > 0$, (H_2) and (H_3) are satisfied. The direct calculation gives

$$B_n^{-1} = \frac{1}{2\,c\,(\lambda - 1)^2} \begin{pmatrix} 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \end{pmatrix} \begin{pmatrix} n^\alpha & \\ & +O(1) \end{pmatrix}$$

whence we obtain

$$\lim_{n \to +\infty} \frac{B_n^{-1}}{|||B_n^{-1}|||_1} = \frac{1}{4} \begin{pmatrix} 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \end{pmatrix} \quad \text{and} \quad \lim_{n \to +\infty} n^{-\alpha}\,|||B_n^{-1}|||_1 = \frac{2}{c\,(\lambda - 1)^2}$$

so (H_4) is satisfied with the 1–norm. The choice $\pi = 0$ is impossible, and we finally find

$$\Gamma_\pi = \frac{\sigma^2}{4} \begin{pmatrix} 1 + \frac{4}{\sigma^2}\,\pi & 1 \\ 1 & 1 + \frac{4}{\sigma^2}\,\pi \end{pmatrix}.$$

2.2.3. *Bivariate case with two nearly unit roots.* Following the same lines, suppose that $p = 2$ and $\mathrm{sp}(A) = \{-1, 1\}$. This situation occurs, for example, when

$$A_n = \begin{pmatrix} (c_2 - c_1)n^{-\alpha} & (1 - c_1\,n^{-\alpha})(1 - c_2\,n^{-\alpha}) \\ 1 & 0 \end{pmatrix}$$

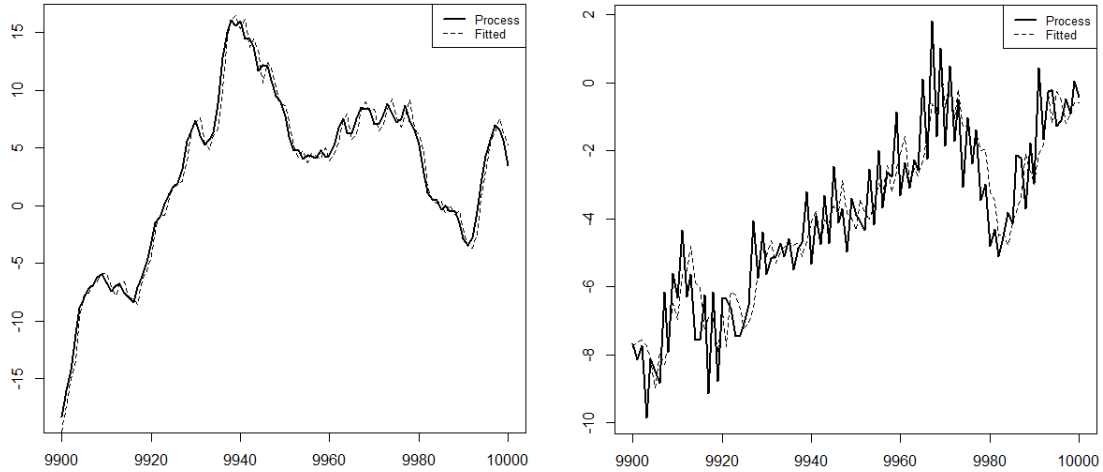


FIGURE 2. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 10^{-5}$ and $\mathcal{N}(0,1)$ innovations. The setting is $\lambda = 0.5$, $c = 0.1$ and $\alpha = 0.32$ on the left, $\lambda = -0.67$, $c = 0.2$ and $\alpha = 0.25$ on the right.

whose eigenvalues are $1 - c_1\, n^{-\alpha}$ and $-1 + c_2\, n^{-\alpha}$. This is illustrated on Figure 3. For $c_1, c_2 > 0$ and $\alpha > 0$, (H_2) and (H_3) are satisfied. The direct calculation gives

$$B_n^{-1} = \frac{1}{8\, c_1\, c_2} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \end{pmatrix} \big(n^{\alpha} + O(1)\big)$$

whence we obtain

$$\lim_{n \to +\infty} \frac{B_n^{-1}}{|||B_n^{-1}|||_1} = \frac{1}{2\,(c_1 + c_2) + 2\,|c_2 - c_1|} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \end{pmatrix}.$$

Moreover,

$$\lim_{n \to +\infty} n^{-\alpha}\, |||B_n^{-1}|||_1 = \frac{(c_1 + c_2) + |c_2 - c_1|}{4\, c_1\, c_2}$$

so (H_4) is satisfied with the 1–norm. The choice $\pi = 0$ is possible and we finally find

$$\Gamma = \frac{\sigma^2}{2\,(c_1 + c_2) + 2\,|c_2 - c_1|} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 \end{pmatrix}.$$

2.3. **Discussion on multiple eigenvalues and conclusion.** As we will see in the proof of Lemma 3.1, the distinct eigenvalues assumption (H_2) is sufficient to reach our results. However, a less stringent formulation of (H_2) could be :

(H_2') *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \to +\infty} A_n = A$$

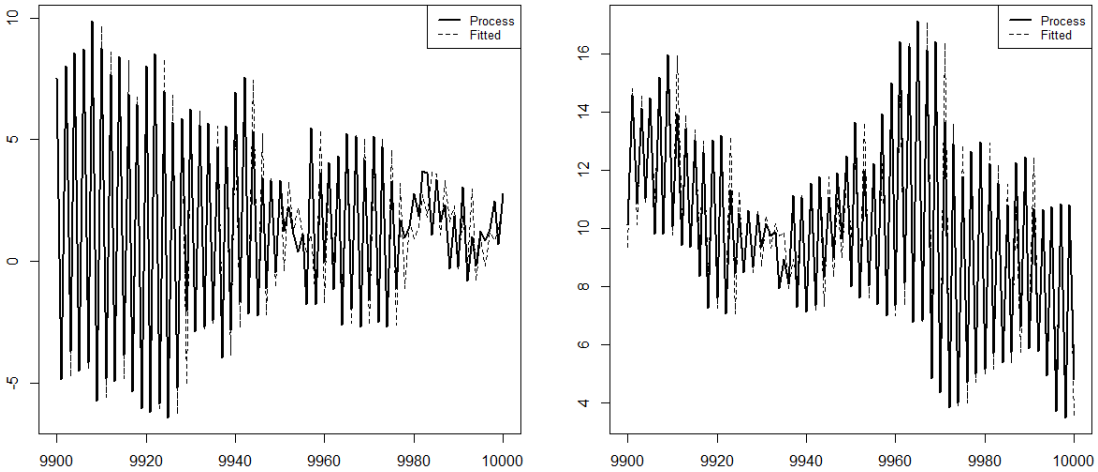


FIGURE 3. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 0$ and $\mathcal{N}(0,1)$ innovations. The setting is $c_1 = 0.1$, $c_2 = 0.2$ and $\alpha = 0.32$ on the left, $c_1 = 0.01$, $c_2 = 0.01$ and $\alpha = 0.25$ on the right.

and the top right element of A is non-zero. In addition, there exists a rank n_0 such that, for all $n > n_0$, A_n is diagonalizable and the change of basis matrix P_n satisfies $|||P_n||| \leqslant C_{st}$ and $|||P_n^{-1}||| \leqslant C_{st}$.

In general, multiple eigenvalues may not falsify our reasonings, except when the multiplicity concerns the eigenvalues whose modulus tends to 1. Indeed, the coefficients of $|||A_n^\ell|||$ may grow faster in that case. Consider the simple bivariate example where

$$A_n^\ell = \begin{pmatrix} a_{11,\,\ell} & a_{12,\,\ell} \\ a_{21,\,\ell} & a_{22,\,\ell} \end{pmatrix} = \begin{pmatrix} \theta_{n,\,1}\, a_{11,\,\ell-1} + \theta_{n,\,2}\, a_{11,\,\ell-2} & \theta_{n,\,2}\, a_{11,\,\ell-1} \\ \theta_{n,\,1}\, a_{21,\,\ell-1} + \theta_{n,\,2}\, a_{21,\,\ell-2} & \theta_{n,\,2}\, a_{21,\,\ell-1} \end{pmatrix}.$$

Then, it is not hard to solve this linear difference equation whose characteristic roots are the eigenvalues of A_n. In case of multiplicity, the top left term takes the form of

$$a_{11,\,\ell} = (c_n + d_n\, \ell)\, \rho^\ell(A_n)$$

and even if $|c_n| \leqslant C_{st}$ and $|d_n| \leqslant C_{st}$ for n large enough, it follows that

$$\sum_{\ell=0}^{+\infty} |||A_n^\ell||| \sim \frac{C_{st}}{(1-\rho(A_n))^2}.$$

That invalidates all our reasonings and, in that case, new approaches are needed to potentially reach the moderate deviations. From our viewpoint, this is the main weakness of the set of hypotheses. As it is already observed in [7], multiple unit roots located at 1 influence the rate of convergence of the OLS. We conjecture that the same phenomenon occurs here and that a larger power should come with $1 - \rho(A_n)$ in the renormalization.

To sum up, this study is a wide generalization of [17] and, although not complete in virtue of the latter remark, it covers most of the MDP issues for the estimation in the stable but nearly unstable case. Large deviations would undoubtedly be a very useful and challenging study to carry out, naturally extending this one. However, to the best of our knowledge, it is not even entirely treated in the stable time-invariant case $\rho(A_n) = \rho(A) < 1$, clearly revealing the complexity of the problem. A complicated but stimulating trail for future

studies could rely on the exponential, and not only polynomial, neighborhood of the unit root. Along the same lines and even if it is of less practical interest, we might as well focus on the explosive side of the unit roots, where new theoretical developments are necessary.

3. Technical proofs

In all the proofs, C_{st} denotes a generic positive constant that is not necessarily identical from one line to another. We will frequently use the fact that $\|\mathrm{vec}(\cdot)\| = |||\cdot|||_F \leqslant C_{st}\,|||\cdot|||$. For asymptotic equivalences, $f_n \asymp g_n$ means that both $f_n = O(g_n)$ and $g_n = O(f_n)$ whereas $f_n \sim g_n$ stands for $\frac{f_n}{g_n} \to 1$.

3.1. **Some linear algebra tools.** Thereafter, we denote by $\lambda_1, \ldots, \lambda_p$ the (distinct) eigenvalues of A and $\lambda_{n,1}, \ldots, \lambda_{n,p}$ those of A_n, in descending order of modulus. We start by establishing two lemmas that will prove to be very useful in what follows.

Lemma 3.1. *Under hypotheses (H_2) and (H_3), as n tends to infinity,*

$$\sum_{\ell=0}^{+\infty} |||A_n^\ell||| \asymp \frac{1}{1-\rho(A_n)} \quad \textit{and} \quad \sum_{\ell=0}^{+\infty} (\ell+1)\,|||A_n^\ell||| \asymp \frac{1}{(1-\rho(A_n))^2}.$$

Proof. The lower bounds are established in Section 1.2, in (1.8) and (1.9). For the upper bounds, fix

$$\delta = \frac{2}{|\lambda_p|}, \quad \epsilon_1 = \frac{1}{2} \min_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left|\frac{1}{\lambda_i} - \frac{1}{\lambda_j}\right| \quad \text{and} \quad \epsilon_2 = 2 \max_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left|\frac{1}{\lambda_i} - \frac{1}{\lambda_j}\right|.$$

According to Thm. 2.4.9.2 of [13], (H_2) implies the existence of a rank $n_0 = n_0(\delta, \epsilon_1, \epsilon_2)$ such that, for all $n > n_0$, the eigenvalues of A_n satisfy

$$0 < \max_{1 \leqslant i \leqslant p} \left|\frac{1}{\lambda_{n,i}}\right| < \delta \tag{3.1}$$

and

$$\epsilon_1 < \min_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left|\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}}\right| < \max_{\substack{1 \leqslant i,j \leqslant p \\ i \neq j}} \left|\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}}\right| < \epsilon_2. \tag{3.2}$$

Let P_n be a change of basis matrix in the diagonalization of A_n. Then, since A_n is a companion matrix, a standard choice would be

$$P_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \frac{1}{\lambda_{n,1}} & \frac{1}{\lambda_{n,2}} & \cdots & \frac{1}{\lambda_{n,p}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\lambda_{n,1}^{p-1}} & \frac{1}{\lambda_{n,2}^{p-1}} & \cdots & \frac{1}{\lambda_{n,p}^{p-1}} \end{pmatrix}. \tag{3.3}$$

This Vandermonde matrix is invertible if and only if $\lambda_{n,i} \neq \lambda_{n,j}$ for all $i \neq j$ (see *e.g.* Sec. 0.9.11 of [13]). In that case, P_n^{-1} is closely related to the Lagrange interpolating polynomials given, for $i \in \{1, \ldots, p\}$, by

$$L_i(X) = \frac{\prod_{j \neq i}(X - \frac{1}{\lambda_{n,j}})}{\prod_{j \neq i}(\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}})}.$$

Precisely, the i–th row of P_n^{-1} contains the coefficients of $L_i(X)$ in the basis $(1, X, \ldots, X^{p-1})$ of $\mathbb{R}_{p-1}[X]$, *i.e.*

$$P_n^{-1} = \left(\frac{p_{n,i,j}}{\prod_{j \neq i} (\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}})} \right)_{1 \leqslant i,j \leqslant p} \tag{3.4}$$

where the relation $\prod_{j \neq i}(X - \frac{1}{\lambda_{n,j}}) = p_{n,i,1} + p_{n,i,2}\, X + \ldots + p_{n,i,p}\, X^{p-1}$ enables to identify each $p_{n,i,j}$. Combining (3.1) and (3.2), it follows that, for all $n > n_0$,

$$|||P_n|||_1 \leqslant p\,(1 + \delta + \ldots + \delta^{p-1}) \leqslant C_{st}.$$

We also have $|||P_n^{-1}|||_1 \leqslant C_{st}$ since $\epsilon_1^{p-1} < \prod_{j \neq i} |\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}}| < \epsilon_2^{p-1}$ and since $p_{n,i,j}$ is a finite combination of sums and products of $\frac{1}{\lambda_{n,1}}, \ldots, \frac{1}{\lambda_{n,p}}$. To sum up, for all $\ell \geqslant 0$ and $n > n_0$,

$$A_n^\ell = P_n\, D_n^\ell\, P_n^{-1} \quad \text{where} \quad D_n = \mathrm{diag}(\lambda_{n,1}, \ldots, \lambda_{n,p}).$$

Consequently,

$$\begin{aligned} |||A_n^\ell||| &= |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + |||P_n\, D_n^\ell\, P_n^{-1}|||\, \mathbb{1}_{\{n > n_0\}} \\ &\leqslant |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + |||P_n|||\, |||P_n^{-1}|||\, |||D_n^\ell|||\, \mathbb{1}_{\{n > n_0\}} \\ &\leqslant |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + C_{st}\, \rho^\ell(A_n)\, \mathbb{1}_{\{n > n_0\}}. \end{aligned} \tag{3.5}$$

It only remains to sum over ℓ and to let n tend to infinity to reach the first result. Similarly,

$$(\ell+1)\, |||A_n^\ell||| \leqslant (\ell+1)\, |||A_n^\ell|||\, \mathbb{1}_{\{n \leqslant n_0\}} + C_{st}\, (\ell+1)\, \rho^\ell(A_n)\, \mathbb{1}_{\{n > n_0\}}$$

so we get the second result by following the same lines. □

Lemma 3.2. *Under hypotheses (H_2) and (H_3), we have the convergence*

$$\lim_{n \to +\infty} A_n^{w_n} = 0$$

for any rate (w_n) satisfying $w_n\,(1 - \rho(A_n)) \to +\infty$.

Proof. Consider the rank n_0 introduced in the proof of Lemma 3.1. Then, according to the inequality (3.5),

$$|||A_n^{w_n}||| \leqslant |||A_n^{w_n}|||\, \mathbb{1}_{\{n \leqslant n_0\}} + C_{st}\, \rho^{\,w_n}(A_n)\, \mathbb{1}_{\{n > n_0\}} \tag{3.6}$$

where the invertible and uniformly bounded matrices P_n and P_n^{-1} are given in (3.3) and (3.4), respectively. We also have

$$\lim_{n \to +\infty} \rho^{\,w_n}(A_n) = \lim_{n \to +\infty} \mathrm{e}^{\,-w_n\,(1-\rho(A_n))} = 0 \tag{3.7}$$

from the hypothesis on (w_n). It remains to let n tend to infinity in the above inequality. □

3.2. **Proofs of the main results.** First of all, it is convenient to express the empirical variance of the process as

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n}(\Phi_{n,k}\,\Phi_{n,k}^T - \Gamma_n) &= \frac{1}{n}\sum_{k=1}^{n} A_n\,\Phi_{n,k-1}\,\Phi_{n,k-1}^T\,A_n^T + \frac{1}{n}\sum_{k=1}^{n} A_n\,\Phi_{n,k-1}\,E_k^T \\ &\quad + \frac{1}{n}\sum_{k=1}^{n} E_k\,\Phi_{n,k-1}^T\,A_n^T + \frac{1}{n}\sum_{k=1}^{n} E_k\,E_k^T - \Gamma_n \end{aligned}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\Delta_{n,\,k} + \frac{1}{n}\sum_{k=1}^{n} A_n\,(\Phi_{n,\,k}\,\Phi_{n,\,k}^T - \Gamma_n)\,A_n^T - \frac{T_n}{n}$$

where the variance Γ_n is given in (1.4),

$$\Delta_n = \frac{1}{n}\sum_{k=1}^{n}\Delta_{n,\,k} \tag{3.8}$$
$$= \frac{1}{n}\sum_{k=1}^{n}(A_n\,\Phi_{n,\,k-1}\,E_k^T + E_k\,\Phi_{n,\,k-1}^T\,A_n^T + E_k\,E_k^T + A_n\,\Gamma_n\,A_n^T - \Gamma_n)$$

and the residual term is

$$T_n = A_n\,(\Phi_{n,\,n}\,\Phi_{n,\,n}^T - \Phi_{n,\,0}\,\Phi_{n,\,0}^T)\,A_n^T.$$

Then, solving this generalized Sylvester equation (Lem. 2.1 of [14]) and considering the invertibility of B_n in (1.7) which is proved at the beggining of Section 1.2, we reach the decomposition

$$\mathrm{vec}\left(\frac{1}{n}\sum_{k=1}^{n}(\Phi_{n,\,k}\,\Phi_{n,\,k}^T - \Gamma_n)\right) = B_n^{-1}\,\mathrm{vec}(\Delta_n) - \frac{B_n^{-1}\,\mathrm{vec}(T_n)}{n}. \tag{3.9}$$

Let us now reason step by step, *via* some intermediate results.

3.2.1. *Exponential moments of the squared initial value.* We recall that, from the causal form (1.3) of the process,

$$\Phi_{n,\,0} = \sum_{\ell=0}^{+\infty} A_n^\ell\,E_{-\ell}.$$

The following result gives an exponential moment for the correctly renormalized squared initial value.

Lemma 3.3. *Under hypothesis (H_1),*

$$\mathbb{E}\left[\exp\left(\frac{\alpha}{L_n^2}\,|||\Phi_{n,\,0}\,\Phi_{n,\,0}^T|||\right)\right] < +\infty$$

where L_n is given in (1.8).

Proof. By Cauchy-Schwarz inequality,

$$|||\Phi_{n,\,0}\,\Phi_{n,\,0}^T||| \leqslant \|\Phi_{n,\,0}\|^2 \leqslant \left(\sum_{\ell=0}^{+\infty}\|A_n^\ell\,E_{-\ell}\|\right)^2$$
$$\leqslant \left(\sum_{\ell=0}^{+\infty}|||A_n^\ell|||^{\frac{1}{2}}\,|||A_n^\ell|||^{\frac{1}{2}}\,\|E_{-\ell}\|\right)^2 \leqslant L_n\sum_{\ell=0}^{+\infty}|||A_n^\ell|||\,\varepsilon_{-\ell}^{\,2}.$$

Moreover, from Jensen's inequality, for all $\lambda > 0$,

$$\exp\left(\frac{\lambda}{L_n}\sum_{\ell=0}^{+\infty}|||A_n^\ell|||\,\varepsilon_{-\ell}^{\,2}\right) \leqslant \frac{1}{L_n}\sum_{\ell=0}^{+\infty}|||A_n^\ell|||\,\mathrm{e}^{\lambda\,\varepsilon_{-\ell}^{\,2}}$$

using $\frac{|||A_n^0|||}{L_n} + \frac{|||A_n^1|||}{L_n} + \ldots = 1$. Taking the expectation and choosing $\lambda = \alpha$ given in (H_1), we deduce that

$$\begin{aligned} \mathbb{E}\Big[\exp\Big(\frac{\alpha}{L_n^2} |||\Phi_{n,0}\,\Phi_{n,0}^T||| \Big) \Big] &\leqslant \frac{1}{L_n} \sum_{\ell=0}^{+\infty} |||A_n^\ell|||\, \mathbb{E}\big[\mathrm{e}^{\alpha\, \varepsilon_{-\ell}^2} \big] \\ &= \mathbb{E}\big[\mathrm{e}^{\alpha\, \varepsilon_1^2} \big] < +\infty. \end{aligned} \tag{3.10}$$

□

3.2.2. *Exponential convergence of the residual term.* The residual term in the decomposition (3.9) is given by

$$R_n = \frac{B_n^{-1}\, \mathrm{vec}(A_n\, (\Phi_{n,n}\,\Phi_{n,n}^T - \Phi_{n,0}\,\Phi_{n,0}^T)\, A_n^T)}{n}. \tag{3.11}$$

Our next objective is to prove the exponential negligibility of this residual.

Lemma 3.4. *Under hypotheses (H_1)–(H_5), for all $r > 0$,*

$$\lim_{n\,\to\,+\infty} \frac{1}{b_n^2} \ln \mathbb{P}\bigg(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\, \|R_n\| \geqslant r \bigg) = -\infty.$$

Proof. First, note that

$$\begin{aligned} \|R_n\| &\leqslant \frac{|||B_n^{-1}|||\, \|\mathrm{vec}(A_n\, (\Phi_{n,n}\,\Phi_{n,n}^T - \Phi_{n,0}\,\Phi_{n,0}^T)\, A_n^T)\|}{n} \\ &\leqslant \frac{C_{st}\, |||B_n^{-1}|||\, |||A_n|||^2\, |||\Phi_{n,n}\,\Phi_{n,n}^T - \Phi_{n,0}\,\Phi_{n,0}^T|||}{n} \\ &\leqslant \frac{C_{st}\, |||B_n^{-1}|||\, |||A_n|||^2\, (|||\Phi_{n,n}\,\Phi_{n,n}^T||| + |||\Phi_{n,0}\,\Phi_{n,0}^T|||)}{n}. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{P}\bigg(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\, \|R_n\| \geqslant r \bigg) &= \mathbb{P}\bigg(\|R_n\| \geqslant \frac{r\, b_n\, (1-\rho(A_n))^{-\frac{3}{2}}}{\sqrt{n}} \bigg) \\ &\leqslant 2\, \mathbb{P}\bigg(|||\Phi_{n,0}\,\Phi_{n,0}^T||| \geqslant \frac{r\, b_n \sqrt{n}\, (1-\rho(A_n))^{-\frac{3}{2}}}{2\, C_{st}\, |||A_n|||^2\, |||B_n^{-1}|||} \bigg) \\ &\leqslant 2\, \mathbb{E}\big[\mathrm{e}^{\alpha\, \varepsilon_1^2} \big] \exp\bigg(- \frac{r\, \alpha\, b_n \sqrt{n}\, (1-\rho(A_n))^{-\frac{3}{2}}}{2\, C_{st}\, |||A_n|||^2\, |||B_n^{-1}|||\, L_n^2} \bigg) \end{aligned}$$

where L_n is given in (1.8), using Markov's inequality, the reasoning in the proof of Lemma 3.3 and the fact that, from the strict stationarity of the process, $\Phi_{n,0}\,\Phi_{n,0}^T$ and $\Phi_{n,n}\,\Phi_{n,n}^T$ share the same distribution. Hence, for a sufficiently large n,

$$\begin{aligned} \frac{1}{b_n^2} \ln \mathbb{P}\bigg(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\, \|R_n\| \geqslant r \bigg) &\leqslant \frac{\ln 2 + \ln \mathbb{E}\big[\mathrm{e}^{\alpha\, \varepsilon_1^2} \big]}{b_n^2} - \frac{r\, \alpha\, \sqrt{n}\, (1-\rho(A_n))^{-\frac{3}{2}}}{2\, C_{st}\, b_n\, |||A_n|||^2\, |||B_n^{-1}|||\, L_n^2} \\ &\leqslant \frac{\ln 2 + \ln \mathbb{E}\big[\mathrm{e}^{\alpha\, \varepsilon_1^2} \big]}{b_n^2} - C_{st}\, \frac{\sqrt{n}\, (1-\rho(A_n))^{\frac{3}{2}}}{b_n} \end{aligned}$$

since $|||B_n^{-1}|||^{\frac{1}{2}} \sim \sqrt{h}\,(1-\rho(A_n))^{-\frac{1}{2}}$ from (H_4), $L_n^2 = O((1-\rho(A_n))^{-2})$ from Lemma 3.1 and since, from (H_2), $|||A_n|||$ converges. Finally, letting n tend to infinity, (H_1) and (H_5) conclude the proof. □

3.2.3. *The truncated sequence.* In what follows, we define the rate

$$m_n = \left\lfloor \left(\frac{1}{1-\rho(A_n)} \right)^{\frac{3+3\eta}{3+2\eta}} \right\rfloor \tag{3.12}$$

and we note from (H_3)–(H_5) that

$$\lim_{n\to+\infty} m_n(1-\rho(A_n)) = +\infty \quad \text{and} \quad \lim_{n\to+\infty} \frac{b_n\,|||B_n^{-1}|||^{\frac{1}{2}}\,m_n^{1+\frac{2\eta}{3}}}{\sqrt{n}} = 0. \tag{3.13}$$

Following the idea of [17], we are going to use m_n as a truncation parameter. Consider

$$\Psi_{n,k} = \sum_{\ell=0}^{m_n-2} A_n^\ell\, E_{k-\ell} \tag{3.14}$$

as an approximation of $\Phi_{n,k}$ in its causal form (1.3). We also define the truncated version of the summands $\Delta_{n,k}$ in (3.8) as

$$\zeta_{n,k} = A_n\,\Psi_{n,k-1}\,E_k^T + E_k\,\Psi_{n,k-1}^T\,A_n^T + E_k\,E_k^T + A_n\,\Gamma_n\,A_n^T - \Gamma_n. \tag{3.15}$$

The process $(B_n^{-1}\,\mathrm{vec}(\zeta_{n,k}))_k$ is strictly stationary and m_n–dependent, according to Def. 6.4.3 of [4]. Let us study some properties of this process.

Lemma 3.5. *Under hypotheses (H_1)–(H_4), we can find a constant $c_\alpha > 0$ such that, for a sufficiently large n,*

$$\mathbb{E}\left[\exp\left(c_\alpha\,|||B_n^{-1}|||^{-1} \sum_{\ell=0}^{w_n} |||A_n^\ell\,E_{-\ell}\,E_1^T||| \right) \right] \leqslant \mathbb{E}\big[\mathrm{e}^{\alpha\,\varepsilon_1^2}\big]$$

for any rate (w_n) satisfying $w_n\,(1-\rho(A_n)) \to +\infty$.

Proof. By Hölder's inequality,

$$\mathbb{E}\left[\exp\left(c_\alpha\,|||B_n^{-1}|||^{-1} \sum_{\ell=0}^{w_n} |||A_n^\ell\,E_{-\ell}\,E_1^T||| \right) \right] \leqslant \mathbb{E}\left[\exp\left(c_\alpha\,|||B_n^{-1}|||^{-1} \sum_{\ell=0}^{w_n} |||A_n^\ell|||\,\varepsilon_1^2 \right) \right].$$

Moreover, for the rank n_0 and the uniformly bounded matrices P_n and P_n^{-1} introduced in the proof of Lemma 3.1,

$$\begin{aligned} \sum_{\ell=0}^{w_n} |||A_n^\ell||| &= \sum_{\ell=0}^{n_0} |||A_n^\ell||| + \sum_{\ell=n_0+1}^{w_n} |||A_n^\ell||| \\ &= \sum_{\ell=0}^{n_0} |||A_n^\ell||| + \sum_{\ell=n_0+1}^{w_n} |||P_n\,D_n^\ell\,P_n^{-1}||| \;\leqslant\; C_{st}\left(1 + \frac{1-\rho^{\,w_n}(A_n)}{1-\rho(A_n)}\right) \end{aligned}$$

as soon as $w_n > n_0$. Thus,

$$|||B_n^{-1}|||^{-1} \sum_{\ell=0}^{w_n} |||A_n^\ell||| \;\leqslant\; C_{st}\left(|||B_n^{-1}|||^{-1} + \frac{1-\rho^{\,w_n}(A_n)}{|||B_n^{-1}|||\,(1-\rho(A_n))} \right).$$

Finally, (H_4), (1.10) and (3.7) lead, for large values of n, to

$$|||B_n^{-1}|||^{-1} \sum_{\ell=0}^{w_n} |||A_n^\ell||| \leqslant C_{st}.$$

It remains to choose $c_\alpha = \frac{\alpha}{C_{st}}$. □

Lemma 3.6. *Under hypotheses (H_2)–(H_4), for all $n \geqslant 1$ and $k \in \{1, \ldots, n\}$,*

$$\mathbb{E}[\mathrm{vec}(\zeta_{n,k})] = 0 \quad \textit{and} \quad \mathbb{C}\mathrm{ov}(\mathrm{vec}(\zeta_{n,k}), \mathrm{vec}(\zeta_{n,j})) = \begin{cases} 0 & \textit{for } k \neq j \\ \Upsilon_n & \textit{for } k = j \end{cases}$$

where the $p^2 \times p^2$ covariance Υ_n can be explicitely built in terms of σ^2, A_n and B_n. In addition,

$$\lim_{n \to +\infty} \frac{B_n^{-1} \, \Upsilon_n \, (B_n^{-1})^T}{|||B_n^{-1}|||^3} = \Upsilon$$

where the non-zero limiting matrix Υ is given in (3.18).

Proof. We will use in what follows K_p and U_p defined in (1.5). Let $\mathcal{F}_k = \sigma(\varepsilon_\ell, \, \ell \leqslant k)$ be the σ–algebra of the events occurring up to time k. Then, it is easy to see that

$$\begin{aligned} \mathbb{E}[\mathrm{vec}(\zeta_{n,k})] &= \mathbb{E}[\, \mathbb{E}[\mathrm{vec}(\zeta_{n,k}) \,|\, \mathcal{F}_{k-1}] \,] \\ &= \sigma^2 \, \mathrm{vec}(K_p) + \mathrm{vec}(A_n \, \Gamma_n \, A_n^T - \Gamma_n) \;=\; 0 \end{aligned}$$

in virtue of (1.6). For $k > j$, by direct calculation,

$$\begin{aligned} \mathbb{E}[\mathrm{vec}(\zeta_{n,k}) \, \mathrm{vec}^T(\zeta_{n,j})] &= \mathbb{E}[\, \mathbb{E}[\mathrm{vec}(\zeta_{n,k}) \, \mathrm{vec}^T(\zeta_{n,j}) \,|\, \mathcal{F}_{k-1}] \,] \\ &= \mathbb{E}[\, (\mathbb{E}[\mathrm{vec}(A_n \, \Psi_{n,k-1} \, E_k^T) + \mathrm{vec}(E_k \, \Psi_{n,k-1}^T \, A_n^T) | \, \mathcal{F}_{k-1}] \\ &\qquad + \sigma^2 \, \mathrm{vec}(K_p) + \mathrm{vec}(A_n \, \Gamma_n \, A_n^T - \Gamma_n)) \, \mathrm{vec}^T(\zeta_{n,j})] \;=\; 0 \end{aligned}$$

and the same is true for $j > k$ since $(\mathbb{E}[\mathrm{vec}(\zeta_{n,k}) \, \mathrm{vec}^T(\zeta_{n,j})])^T = \mathbb{E}[\mathrm{vec}(\zeta_{n,j}) \, \mathrm{vec}^T(\zeta_{n,k})] = 0$. Now for $k = j$, a tedious but straightforward calculation leads to

$$\begin{aligned} \mathbb{E}[\mathrm{vec}(\zeta_{n,k}) \, \mathrm{vec}^T(\zeta_{n,k})] &= \sigma^2 \, K_p \otimes (A_n \, \mathbb{E}[\Psi_{n,k-1} \Psi_{n,k-1}^T] \, A_n^T) \\ &\quad + \sigma^2 \, U_p \otimes (A_n \, \mathbb{E}[\Psi_{n,k-1} \Psi_{n,k-1}^T] \, A_n^T) \otimes U_p^T \\ &\quad + \sigma^2 \, U_p^T \otimes (A_n \, \mathbb{E}[\Psi_{n,k-1} \Psi_{n,k-1}^T] \, A_n^T) \otimes U_p \\ &\quad + \sigma^2 \, (A_n \, \mathbb{E}[\Psi_{n,k-1} \Psi_{n,k-1}^T] \, A_n^T) \otimes K_p \\ &\quad + (\tau^4 - \sigma^4) \, \mathrm{vec}(K_p) \, \mathrm{vec}^T(K_p) \;=\; \Upsilon_n. \end{aligned} \tag{3.16}$$

To give an explicit expression of Υ_n, it suffices to observe that the truncated expression (3.14) has a variance given by

$$\Gamma_{n,m_n} = \mathbb{E}[\Psi_{n,k-1} \Psi_{n,k-1}^T] = \sigma^2 \sum_{\ell=0}^{m_n-2} A_n^\ell \, K_p \, (A_n^T)^\ell$$

so that

$$\begin{aligned} \mathrm{vec}(\Gamma_{n,m_n}) &= \sigma^2 \sum_{\ell=0}^{m_n-2} (A_n \otimes A_n)^\ell \, \mathrm{vec}(K_p) \\ &= \sigma^2 \, B_n^{-1} \, (I_{p^2} - (A_n \otimes A_n)^{m_n-1}) \, \mathrm{vec}(K_p). \end{aligned}$$

Let us now look at the asymptotic behavior of Υ_n correctly renormalized. First, we have the convergence

$$\lim_{n\to+\infty} (A_n \otimes A_n)^{m_n-1} = 0$$

coming from the identity $(A_n \otimes A_n)^{m_n-1} = A_n^{m_n-1} \otimes A_n^{m_n-1}$ and Lemma 3.2. Together with (H_4), this implies

$$\lim_{n\to+\infty} \frac{\mathrm{vec}(\Gamma_{n,m_n})}{|||B_n^{-1}|||} = \sigma^2\, H\, \mathrm{vec}(K_p).$$

In the end of the proof, we call vec^{-1} the vectorization inverse operator (namely, in our context, the reconstruction of a $p \times p$ matrix from its vectorization of size p^2). Then,

$$\lim_{n\to+\infty} \frac{\Gamma_{n,m_n}}{|||B_n^{-1}|||} = \sigma^2\, \mathrm{vec}^{-1}(H\, \mathrm{vec}(K_p)) = \Gamma. \tag{3.17}$$

Combining (3.16) with (3.17) and (H_4), we have

$$\Upsilon = \sigma^2\, H\, (K_p \otimes \Gamma^A + U_p \otimes \Gamma^A \otimes U_p^T + U_p^T \otimes \Gamma^A \otimes U_p + \Gamma^A \otimes K_p)\, H^T \tag{3.18}$$

where $\Gamma^A = A\,\Gamma\,A^T$. □

Remark. As a by-product, we also obtain, following the same lines,

$$\lim_{n\to+\infty} \frac{\Gamma_n}{|||B_n^{-1}|||} = \Gamma$$

where Γ_n is given in (1.4), which proves (1.11). The variance Γ_{n,m_n} defined above may be seen as the truncated version of Γ_n.

3.2.4. *The remainder of the truncation.* We denote by

$$\Lambda_n = \frac{1}{n}\sum_{k=1}^{n} (A_n\,(\Phi_{n,k-1} - \Psi_{n,k-1})\,E_k^T + E_k\,(\Phi_{n,k-1} - \Psi_{n,k-1})^T\,A_n^T) \tag{3.19}$$

the remainder of the truncation of Δ_n in (3.8) made *via* (3.15). Our last preliminary objective is to establish the following lemma.

Lemma 3.7. *Under hypotheses (H_1)–(H_5), for all $r > 0$,*

$$\lim_{n\to+\infty} \frac{1}{b_n^2} \ln \mathbb{P}\left(\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\, \|B_n^{-1}\,\mathrm{vec}(\Lambda_n)\| \geqslant r \right) = -\infty.$$

Proof. Clearly, both terms in the definition of (3.19) are similar and we will only work on the first one. From the causal expression (1.3) and the truncation (3.14), we note that

$$\begin{aligned} \sum_{k=1}^{n} A_n\,(\Phi_{n,k-1} - \Psi_{n,k-1})\,E_k^T &= \sum_{k=1}^{n} \sum_{\ell=m_n-1}^{+\infty} A_n^{\ell+1}\,E_{k-1-\ell}\,E_k^T \\ &= A_n^{m_n} \sum_{\ell=0}^{+\infty} A_n^{\ell} \sum_{k=1}^{n} E_{k-\ell-m_n}\,E_k^T. \end{aligned}$$

Thus, with M_n given in (1.9) and applying Lem. 17 of [15] under (H_1),

$$\mathbb{P}\left(\frac{1}{n} \left|\left|\left| \sum_{k=1}^{n} A_n\,(\Phi_{n,k-1} - \Psi_{n,k-1})\,E_k^T \right|\right|\right| \geqslant r\,\frac{b_n}{\sqrt{n}}\,|||B_n^{-1}|||^{\frac{1}{2}} \right)$$

$$
\begin{aligned}
&\leqslant \mathbb{P}\Big(\sum_{\ell=0}^{+\infty} |||A_n^\ell||| \, \Big| \sum_{k=1}^{n} \varepsilon_{k-\ell-m_n} \, \varepsilon_k \Big| \geqslant \sum_{\ell=0}^{+\infty} (\ell+1) \, |||A_n^\ell||| \, \frac{r \, b_n \sqrt{n} \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}|||} \Big) \\
&\leqslant \sum_{\ell=0}^{+\infty} \mathbb{P}\Big(\max_{1 \leqslant j \leqslant n} \Big| \sum_{k=1}^{j} \varepsilon_{k-\ell-m_n} \, \varepsilon_k \Big| \geqslant \frac{r \, (\ell+1) \, b_n \sqrt{n} \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}|||} \Big) \\
&\leqslant C_{st} \sum_{\ell=0}^{+\infty} \exp\Big(- \frac{r^2 \, b_n^2 \, n \, t_{n,\ell}^2}{\alpha_0 \, n + \beta_0 \, r \, t_{n,\ell} \, b_n \sqrt{n}} \Big) \qquad (3.20)
\end{aligned}
$$

for some $\alpha_0 > 0$ and $\beta_0 > 0$, where

$$
t_{n,\ell} = \frac{(\ell+1) \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}|||}.
$$

Our choice of m_n in (1.9), the properties of Lemma 3.1, (3.6) and our hypotheses on the rates of convergence lead, for n large enough, to

$$
|||B_n^{-1}|||^{-\frac{1}{2}} \, M_n \, |||A_n^{m_n}||| \; \leqslant \; C_{st} \, (1 - \rho(A_n))^{-\frac{3}{2}} \, \rho^{\, m_n}(A_n) \; \longrightarrow \; 0
$$

and obviously $t_{n,\ell} \to +\infty$. Hence, like in formula (3.11) of [17], there are some constants $\alpha_0' > 0$ and $\beta_0' > 0$ such that, for all $\ell \geqslant 0$ and large values of n,

$$
\begin{aligned}
\frac{r^2 \, n \, b_n^2 \, t_{n,\ell}^2}{\alpha_0 \, n + \beta_0 \, r \, t_{n,\ell} \, b_n \sqrt{n}} &= \frac{r^2 \, (\ell+1) \, b_n^2 \, t_{n,\ell}}{\alpha_0 \, |||B_n^{-1}|||^{-\frac{1}{2}} \, M_n \, |||A_n^{m_n}||| + r \, \beta_0 \, (\ell+1) \, \frac{b_n}{\sqrt{n}}} \\
&\geqslant b_n^2 \, t_{n,\ell} \, \frac{r^2}{\alpha_0' + r \, \beta_0'}.
\end{aligned}
$$

Going back to (3.20),

$$
\begin{aligned}
\sum_{\ell=0}^{+\infty} \exp\Big(- \frac{r^2 \, b_n^2 \, n \, t_{n,\ell}^2}{\alpha_0 \, n + \beta_0 \, r \, t_{n,\ell} \, b_n \sqrt{n}} \Big) &\leqslant \sum_{\ell=0}^{+\infty} \exp\Big(- b_n^2 \, t_{n,\ell} \, \frac{r^2}{\alpha_0' + r \, \beta_0'} \Big) \\
&= \frac{\mathrm{e}^{-V_n}}{1 - \mathrm{e}^{-V_n}}
\end{aligned}
$$

where, for convenience, we note

$$
V_n = \frac{r^2 \, b_n^2 \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}||| \, (\alpha_0' + r \, \beta_0')} \; \longrightarrow \; +\infty.
$$

To sum up,

$$
\begin{aligned}
&\frac{1}{b_n^2} \ln \mathbb{P}\Big(\Big|\Big| \sum_{k=1}^{n} A_n \, (\Phi_{n,k-1} - \Psi_{n,k-1}) \, E_k^T \Big|\Big| \geqslant r \, b_n \sqrt{n} \, |||B_n^{-1}|||^{\frac{1}{2}} \Big) \\
&\qquad \leqslant \frac{C_{st} - \ln(1 - \mathrm{e}^{-V_n})}{b_n^2} - \frac{V_n}{b_n^2} \\
&\qquad \leqslant \frac{C_{st} - \ln(1 - \mathrm{e}^{-V_n})}{b_n^2} - \frac{r^2 \, |||B_n^{-1}|||^{\frac{1}{2}}}{M_n \, |||A_n^{m_n}||| \, (\alpha_0' + r \, \beta_0')} \; \longrightarrow \; -\infty.
\end{aligned}
$$

This is clearly sufficient to finish the proof since, from (H_4),

$$\begin{aligned}\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\,\|B_n^{-1}\,\mathrm{vec}(\Lambda_n)\| &\leqslant C_{st}\,\frac{\sqrt{n}}{b_n}\,|||B_n^{-1}|||^{-\frac{1}{2}}\,\|\mathrm{vec}(\Lambda_n)\| \\ &\leqslant C_{st}\,\frac{\sqrt{n}}{b_n}\,|||B_n^{-1}|||^{-\frac{1}{2}}\,|||\Lambda_n|||\end{aligned}$$

for n large enough. □

We are now ready to prove Theorem 2.1 and Corollary 2.2.

3.2.5. *Proof of Theorem 2.1.* All the technical results of the previous sections are now going to be concretely used. Consider the sequence

$$\xi_{n,\,k} = \frac{B_n^{-1}\,\mathrm{vec}(\zeta_{n,\,k})}{|||B_n^{-1}|||^{\frac{3}{2}}} \tag{3.21}$$

where $\zeta_{n,\,k}$ is given in (3.15). The process $(\xi_{n,\,k})_k$ is also strictly stationary and m_n–dependent. Like in [18] or [17, suppl. mat.], let us extract an independent sequence from this process. For $j \in \{1, \ldots, j_n\}$, define

$$\xi'_{n,\,j} = \xi_{n,\,(j-1)m_n+1} + \ldots + \xi_{n,\,jm_n}$$

where $j_n = \lfloor \frac{n}{m_n} \rfloor$ and where (m_n) and its properties are given in (3.12). Then, $(\xi'_{n,\,j})_j$ is strictly stationary and 1–dependent. Next, for $t \in \{1, \ldots, t_n\}$, define

$$\xi''_{n,\,t} = \xi'_{n,\,(t-1)u_n+1} + \ldots + \xi'_{n,\,tu_n-1}$$

where $t_n = \lfloor \frac{j_n}{u_n} \rfloor$ and (u_n) is another rate satisfying

$$\lim_{n\,\to\,+\infty} u_n = +\infty \quad \text{and} \quad \lim_{n\,\to\,+\infty} \frac{b_n\,|||B_n^{-1}|||^{\frac{1}{2}}\,(m_n\,u_n)^{1+\frac{2\eta}{3}}}{\sqrt{n}} = 0. \tag{3.22}$$

To be convinced that such a rate exists, one can use (3.13) and the fact that $|\ln f_n| \to +\infty$ and $f_n\,|\ln f_n|^a \to 0$ when $f_n \to 0$. The process $(\xi''_{n,\,t})_t$ is now i.i.d. and the rates satisfy

$$\lim_{n\,\to\,+\infty} \frac{t_n\,u_n\,m_n}{n} = 1. \tag{3.23}$$

The reasoning of [17, suppl. mat.] does not suit us, so we need to reformulate the establishment of the MDP. First, by a Taylor-Lagrange expansion,

$$\exp\Big(\Big\langle \lambda,\, \frac{b_n}{\sqrt{n}}\,\xi''_{n,\,1}\Big\rangle\Big) = 1 + \frac{b_n}{\sqrt{n}}\,\langle \lambda,\, \xi''_{n,\,1}\rangle + \frac{b_n^2}{2\,n}\,\langle \lambda,\, \xi''_{n,\,1}\rangle^2 + \frac{b_n^3}{6\,n^{\frac{3}{2}}}\,\langle \lambda,\, \xi''_{n,\,1}\rangle^3\,\mathrm{e}^{\,\nu_n} \tag{3.24}$$

in which the remainder term satisfies, for any $\alpha > 0$,

$$\begin{aligned}\mathrm{e}^{\,\alpha\,\nu_n} \;<\; \exp\Big(\frac{\alpha\,b_n}{\sqrt{n}}\,|\langle \lambda,\, \xi''_{n,\,1}\rangle|\Big) &\leqslant \exp\Big(\frac{\alpha\,b_n}{\sqrt{n}}\,\|\lambda\|\,\sum_{\ell=1}^{m_n u_n}\|\xi_{n,\,\ell}\|\Big) \\ &\leqslant \exp\Big(C_{st}\,\frac{b_n}{\sqrt{n}}\,|||B_n^{-1}|||^{-\frac{1}{2}}\,\sum_{\ell=1}^{m_n u_n}|||\zeta_{n,\,\ell}|||\Big).\end{aligned}$$

Now, the random variables $|||\zeta_{n,\ell}|||$ sharing the same distribution for all $\ell \geqslant 0$, it follows from Hölder's inequality that,

$$\begin{aligned}\mathbb{E}\big[\mathrm{e}^{\alpha\,\nu_n}\big] &< \mathbb{E}\Big[\exp\Big(C_{st}\,\frac{b_n\,m_n\,u_n}{\sqrt{n}}\,|||B_n^{-1}|||^{-\frac{1}{2}}\,|||\zeta_{n,1}|||\Big)\Big]\\ &= \mathbb{E}\Big[\exp\Big(C_{st}\,\frac{b_n\,|||B_n^{-1}|||^{\frac{1}{2}}\,m_n\,u_n}{\sqrt{n}}\,|||B_n^{-1}|||^{-1}\,|||\zeta_{n,1}|||\Big)\Big] < +\infty \qquad (3.25)\end{aligned}$$

for n large enough, using Lemma 3.5 with $m_n\,(1-\rho(A_n)) \to +\infty$ stemming from (3.13), the convergence of $|||A_n|||$, (H_1) and treating all the terms of (3.15) similarly. Taking the expectation in (3.24) and exploiting the independence of the zero-mean process $(\xi''_{n,t})_t$, we obtain the decomposition

$$\begin{aligned}\frac{1}{b_n^2}\ln\mathbb{E}\Big[\exp\Big(\Big\langle\lambda,\,\frac{b_n}{\sqrt{n}}\sum_{\ell=1}^{n}\xi_{n,\ell}\Big\rangle\Big)\Big] &\sim \frac{t_n}{b_n^2}\ln\mathbb{E}\Big[\exp\Big(\Big\langle\lambda,\,\frac{b_n}{\sqrt{n}}\,\xi''_{n,1}\Big\rangle\Big)\Big]\\ &= \frac{t_n}{2\,n}\,\mathbb{E}\big[\langle\lambda,\,\xi''_{n,1}\rangle^2\big] + O\Big(\frac{t_n\,b_n}{6\,n^{\frac{3}{2}}}\,\big|\mathbb{E}\big[\langle\lambda,\,\xi''_{n,1}\rangle^3\,\mathrm{e}^{\,\nu_n}\big]\big|\Big) \qquad (3.26)\end{aligned}$$

for we can see, as it is done in [18], that the residual term

$$\tau_n = \sum_{\ell=1}^{n}\xi_{n,\ell} - \sum_{\ell=1}^{t_n}\xi''_{n,\ell}$$

plays a negligible role in comparison to the main one. To eliminate the third-order term, we first look at the fourth-order moment of $\langle\lambda,\,\xi''_{n,1}\rangle$, that is

$$\mathbb{E}\big[\langle\lambda,\,\xi''_{n,1}\rangle^4\big] \leqslant \frac{C_{st}\,\|\lambda\|^4}{|||B_n^{-1}|||^2}\,\mathbb{E}\Big[\Big|\Big|\Big|\sum_{\ell=1}^{m_n u_n}\zeta_{n,\ell}\Big|\Big|\Big|^4\Big].$$

A long but standard calculation shows that

$$\begin{aligned}\mathbb{E}\Big[\Big|\Big|\Big|A_n\sum_{\ell=1}^{m_n u_n}\Psi_{n,\ell-1}\,E_\ell^T\Big|\Big|\Big|^4\Big] &\leqslant C_{st}\,\mathbb{E}\Big[\Big\|\sum_{\ell=1}^{m_n u_n}\Psi_{n,\ell-1}\,\varepsilon_\ell\Big\|^4\Big]\\ &= O\big((m_n\,u_n\,|||B_n^{-1}|||)^2\big)\end{aligned}$$

as n tends to infinity. This result is reached using the strict stationarity of the process, the explicit expression of $X_{n,0}^4$ in terms of A_n^ℓ, the inequality (3.6) and, finally, using (H_4) giving the equivalence between $(1-\rho(A_n))^{-2}$ and $C_{st}\,|||B_n^{-1}|||^2$. So,

$$\mathbb{E}\big[\langle\lambda,\,\xi''_{n,1}\rangle^4\big] = O(m_n^2\,u_n^2).$$

By Lyapunov's inequality,

$$\mathbb{E}\big[|\langle\lambda,\,\xi''_{n,1}\rangle|^{3+\delta}\big] \leqslant \big(\mathbb{E}\big[\langle\lambda,\,\xi''_{n,1}\rangle^4\big]\big)^{\frac{3+\delta}{4}} = O\Big((m_n\,u_n)^{\frac{3+\delta}{2}}\Big)$$

for a small $\delta > 0$. Now, combining this result with (3.25) and Hölder's inequality, for sufficiently large values of n,

$$\frac{t_n\,b_n}{n^{\frac{3}{2}}}\,\mathbb{E}\big[|\langle\lambda,\,\xi''_{n,1}\rangle^3\,\mathrm{e}^{\,\nu_n}|\big] \leqslant \frac{t_n\,b_n}{n^{\frac{3}{2}}}\,\big(\mathbb{E}\big[|\langle\lambda,\,\xi''_{n,1}\rangle|^{3+\delta}\big]\big)^{\frac{3}{3+\delta}}\,\big(\mathbb{E}\big[\mathrm{e}^{\frac{3+\delta}{\delta}\,\nu_n}\big]\big)^{\frac{\delta}{3+\delta}}$$

$$\leqslant \; C_{st}\,\frac{t_n\,b_n}{n^{\frac{3}{2}}}\,(m_n\,u_n)^{\frac{3}{2}} \;\longrightarrow\; 0 \tag{3.27}$$

by (3.25), (3.23) and the properties in (3.22). The second-order term in (3.26) satisfies

$$\begin{aligned}\frac{t_n}{2\,n}\,\mathbb{E}\big[\langle\lambda,\,\xi''_{n,\,1}\rangle^2\big] \;=\; \frac{t_n}{2\,n}\,\lambda^T\,\mathbb{V}(\xi''_{n,\,1})\,\lambda \;&=\; \frac{t_n\,u_n\,m_n}{2\,n\,|||B_n^{-1}|||^3}\,\lambda^T\,B_n^{-1}\,\mathbb{V}(\mathrm{vec}(\zeta_{n,\,1}))(B_n^{-1})^T\lambda\\ &=\; \frac{t_n\,u_n\,m_n}{2\,n}\,\lambda^T\,\frac{B_n^{-1}\,\Upsilon_n\,(B_n^{-1})^T}{|||B_n^{-1}|||^{\,3}}\,\lambda\\ &\longrightarrow\; \frac{1}{2}\,\langle\lambda,\Upsilon\,\lambda\rangle\end{aligned} \tag{3.28}$$

where we used (3.23) and the results of Lemma 3.6. The combination of (3.26), (3.27) and (3.28) together with the Gärtner-Ellis theorem (see *e.g.* Sec. 2.3 of [8]) shows that the sequence

$$\left(\frac{1}{b_n\,\sqrt{n}}\sum_{\ell=1}^{n}\xi_{n,\,\ell}\right)_{n\,\geqslant\,1}$$

satisfies an LDP with speed (b_n^2) and rate function given by the Fenchel-Legendre transform of the above logarithmic moment generating function, *i.e.*

$$I(x) = \sup_{\lambda\,\in\,\mathbb{R}^{p^2}}\Big\{\langle\lambda,x\rangle - \frac{1}{2}\,\langle\lambda,\Upsilon\,\lambda\rangle\Big\}.$$

Note that, due to its particular structure, Υ is only non-negative definite as soon as $p > 1$ (by way of example, its last row and column are zero). In that case (see *e.g.* Ex. 1.1.4 of [12], page 212), the explicit expression of this quadratic rate function, strictly convex on its relative interior, is

$$I(x) = \begin{cases}\frac{1}{2}\,\langle x,\Upsilon^{\,\dagger}\,x\rangle & \text{for } x\in\mathrm{Im}(\Upsilon)\\ +\infty & \text{otherwise.}\end{cases}$$

After the truncation introduced in (3.14), the decomposition (3.9) can be rewritten as

$$\begin{aligned}\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\,\mathrm{vec}\left(\frac{1}{n}\sum_{k=1}^{n}(\Phi_{n,\,k}\,\Phi_{n,\,k}^T-\Gamma_n)\right) \;&=\; \frac{(1-\rho(A_n))^{\frac{3}{2}}\,|||B_n^{-1}|||^{\frac{3}{2}}}{b_n\,\sqrt{n}}\,\sum_{k=1}^{n}\xi_{n,\,k}\\ &\quad+\;\frac{\sqrt{n}\,(1-\rho(A_n))^{\frac{3}{2}}}{b_n}\,R_n^*\end{aligned}$$

where, in the remainder term $R_n^* = B_n^{-1}\,\mathrm{vec}(\Lambda_n) - R_n$, the residual of the truncation is given in (3.19) and the main residual R_n is given in (3.11). Lemma 3.4 and Lemma 3.7 show that the first term in the right-hand is an exponentially good approximation of the left-hand side and that, as a consequence, they share the same LDP (see Def. 4.2.10 and Thm. 4.2.13 of [8]). The contraction principle (see Thm. 4.2.1 of [8]) enables to compute the rate function associated with the LDP, namely

$$I_\Gamma(x) = I\big(h^{-\frac{3}{2}}\,x\big) = \begin{cases}\frac{1}{2\,h^3}\,\langle x,\Upsilon^{\,\dagger}\,x\rangle & \text{for } x\in\mathrm{Im}(\Upsilon)\\ +\infty & \text{otherwise}\end{cases} \tag{3.29}$$

where the limiting value $h > 0$ comes from (H_4). □

21

3.2.6. *Proof of Corollary 2.2.* Using (2.2) and (2.3),

$$\begin{aligned}\frac{\sqrt{n}}{b_n\,(1-\rho(A_n))^{\frac{1}{2}}}\left(\widehat{\theta}_n^{\pi}-\theta_n^{\pi}\right) &= \frac{\sqrt{n}\,(S_{n-1}^{\pi})^{-1}}{b_n\,(1-\rho(A_n))^{\frac{1}{2}}}\sum_{k=1}^{n}\Phi_{n,\,k-1}\,\varepsilon_k\\ &= \frac{n\,|||B_n^{-1}|||\,(S_{n-1}^{\pi})^{-1}}{b_n\,\sqrt{n}\,|||B_n^{-1}|||^{\frac{1}{2}}\,(1-\rho(A_n))^{\frac{1}{2}}\,|||B_n^{-1}|||^{\frac{1}{2}}}\sum_{k=1}^{n}\Phi_{n,\,k-1}\,\varepsilon_k.\end{aligned}$$

Our objective is first to prove that, for all $r>0$,

$$\lim_{n\to+\infty}\frac{1}{b_n^2}\ln\mathbb{P}\left(\left|\left|\left|n\,|||B_n^{-1}|||\,(S_{n-1}^{\pi})^{-1}-\Gamma_{\pi}^{-1}\right|\right|\right|\geqslant r\right)=-\infty \tag{3.30}$$

where Γ_π is the invertible penalized variance (2.3), and then to establish an LDP for the sequence

$$\left(\frac{1}{b_n\,\sqrt{n}\,|||B_n^{-1}|||^{\frac{1}{2}}}\sum_{k=1}^{n}\Phi_{n,\,k-1}\,\varepsilon_k\right)_{n\geqslant 1} \tag{3.31}$$

in order to obtain the announced result, *via* the contraction principle (Thm. 4.2.1 of [8]). On the one hand, we know from Theorem 2.1 and (3.29) that

$$\begin{aligned}\frac{1}{b_n^2}\ln\mathbb{P}\left(\left|\left|\left|\frac{S_{n-1}}{n\,|||B_n^{-1}|||}-\frac{\Gamma_n}{|||B_n^{-1}|||}\right|\right|\right|\geqslant r\right) &= \frac{1}{b_n^2}\ln\mathbb{P}\left(\frac{\sqrt{n}}{b_n\,|||B_n^{-1}|||^{\frac{3}{2}}}\left|\left|\left|\frac{S_{n-1}}{n}-\Gamma_n\right|\right|\right|\geqslant r_n\right)\\ &\longrightarrow -\infty = -\lim_{\|x\|\to+\infty} I_\Gamma(x)\end{aligned}$$

since, by (H_4) and (H_5),

$$r_n=\frac{r\,\sqrt{n}}{b_n\,|||B_n^{-1}|||^{\frac{1}{2}}}\ \longrightarrow\ +\infty$$

and $(1-\rho(A_n))^{\frac{3}{2}}\sim h^{\frac{3}{2}}\,|||B_n^{-1}|||^{-\frac{3}{2}}$. So,

$$\lim_{n\to+\infty}\frac{1}{b_n^2}\ln\mathbb{P}\left(\left|\left|\left|\frac{S_{n-1}^{\pi}}{n\,|||B_n^{-1}|||}-\Gamma_n^{\pi}\right|\right|\right|\geqslant r\right)=-\infty\quad\text{for}\quad\Gamma_n^{\pi}=\frac{\Gamma_n}{|||B_n^{-1}|||}+\pi\,I_p.$$

It is also clear that

$$\left\{\left|\left|\left|\frac{S_{n-1}^{\pi}}{n\,|||B_n^{-1}|||}-\Gamma_{\pi}\right|\right|\right|\geqslant r\right\}\ \subset\ \left\{\left|\left|\left|\frac{S_{n-1}^{\pi}}{n\,|||B_n^{-1}|||}-\Gamma_n^{\pi}\right|\right|\right|\geqslant\frac{r}{2}\right\}\cup\left\{|||\Gamma_n^{\pi}-\Gamma_{\pi}|||\geqslant\frac{r}{2}\right\}$$

and (1.11) shows that the second event in the right-hand side becomes impossible when n increases. Hence, from the reasoning above,

$$\lim_{n\to+\infty}\frac{1}{b_n^2}\ln\mathbb{P}\left(\left|\left|\left|\frac{S_{n-1}^{\pi}}{n\,|||B_n^{-1}|||}-\Gamma_{\pi}\right|\right|\right|\geqslant r\right)=-\infty.$$

Now we shall use Lem. 2 of [21] to get (3.30).

On the other hand, all the work consisting in proving that the sequence (3.31) satisfies an LDP with speed (b_n^2) has already been done in the proof of Theorem 2.1. Indeed, *via* the truncation (3.14),

$$\frac{1}{b_n\,\sqrt{n}\,|||B_n^{-1}|||^{\frac{1}{2}}}\sum_{k=1}^{n}\Psi_{n,\,k-1}\,\varepsilon_k\ =\ \frac{1}{b_n\,\sqrt{n}\,|||B_n^{-1}|||^{\frac{1}{2}}}\sum_{k=1}^{n}\sum_{\ell=0}^{m_n-2}A_n^{\ell}\,E_{k-\ell-1}\,\varepsilon_k$$

$$= \frac{1}{b_n \sqrt{n}} \sum_{k=1}^{n} Z_{n,k}$$

where the process $(Z_{n,k})_k$ forms a strictly stationary and m_n–dependent sequence. However, apart from the renormalization, this is precisely the first column of the first term of (3.15). Thus, the calculations are similar and we find, like in Lemma 3.6,

$$\mathbb{V}(Z_{n,1}) = \frac{\sigma^2 \, \Gamma_{n,m_n}}{|||B_n^{-1}|||}.$$

In that case, from the convergence (3.17) and the previous proof, the rate function associated with the LDP is given by

$$J(x) = \sup_{\lambda \in \mathbb{R}^p} \left\{ \langle \lambda, x \rangle - \frac{\sigma^2}{2} \langle \lambda, \Gamma \, \lambda \rangle \right\} = \begin{cases} \frac{1}{2\sigma^2} \langle x, \Gamma^\dagger \, x \rangle & \text{for } x \in \mathrm{Im}(\Gamma) \\ +\infty & \text{otherwise.} \end{cases}$$

The exponential negligibility of the remainder of the truncation is obtained by following the lines of Lemma 3.7. The contraction principle enables to compute the rate function associated with the LDP, namely

$$I_\theta(x) = J\big(\Gamma_\pi \sqrt{h}\, x\big) = \begin{cases} \frac{h}{2\sigma^2} \langle x, \Gamma_\pi \, \Gamma^\dagger \, \Gamma_\pi \, x \rangle & \text{for } x \in \mathrm{Im}(\Gamma_\pi^{-1} \, \Gamma) \\ +\infty & \text{otherwise} \end{cases} \tag{3.32}$$

where the exponential convergence (3.30) has been combined to the LDP established on the sequence (3.31). □

Acknowledgements. The author thanks the associate editor and the two anonymous reviewers for the numerous comments and suggestions that clearly helped to improve the paper. He also thanks R. Garbit for the constructive discussion about the link between Vandermonde matrices and Lagrange polynomials. This work is part of the research program PANORisk of the Région Pays de la Loire.

References

[1] Bercu, B. On large deviations in the Gaussian autoregressive process: stable, unstable and explosive cases. *Bernoulli. 7* (2001), 299–316.

[2] Bercu, B., Gamboa, F., and Rouault, A. Large deviations for quadratic forms of stationary Gaussian processes. *Stoch. Proc. Appl. 71* (1997), 75–90.

[3] Bitseki Penda, V., Djellout, H., and Proïa, F. Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. *ESAIM Probab. Stat. 18* (2014), 308–331.

[4] Brockwell, P. J., and Davis, R. A. *Time series: Theory and Methods (Second Edition)*. Springer Series in Statistics. Springer, New York, 1991.

[5] Buchmann, B., and Chan, N. H. Unified asymptotic theory for nearly unstable AR(p) processes. *Stoch. Proc. Appl. 123* (2013), 952–985.

[6] Chan, N. H., and Wei, C. Z. Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Stat. 15* (1987), 1050–1063.

[7] Chan, N. H., and Wei, C. Z. Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist. 16* (1988), 367–401.

[8] Dembo, A., and Zeitouni, O. *Large Deviations Techniques and Applications (Second Edition)*, vol. 38 of *Applications of Mathematics*. Springer, 1998.

[9] Djellout, H., Guillin, A., and Wu, L. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. I. H. Poincaré. 42* (2006), 393–416.

[10] Donsker, M. D., and Varadhan, S. R. S. Large deviations for stationary Gaussian processes. *Comm. Math. Phys. 97* (1985), 187–210.

[11] Duflo, M. *Random iterative models*. Applications of Mathematics (vol. 34), New York. Springer-Verlag, Berlin, 1997.
[12] Hiriart-Urruty, J. B., and Lemaréchal, C. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer, 2012.
[13] Horn, R. A., and Johnson, C. R. *Matrix Analysis (Second Edition)*. Cambridge University Press, Cambridge, New-York, 2012.
[14] Jiang, T., and Wei, M. On solutions of the matrix equations $X - AXB = C$ and $X - A\bar{X}B = C$. *Linear Algebra Appl. 367* (2003), 225–233.
[15] Mas, A., and Menneteau, L. Large and moderate deviations for infinite-dimensional autoregressive processes. *J. Multivariate Anal. 87* (2003), 241–260.
[16] Miao, Y., and Shen, S. Moderate deviation principle for autoregressive processes. *J. Multivariate Anal. 100* (2009), 1952–1961.
[17] Miao, Y., Wang, Y., and Yang, G. Moderate deviation principles for empirical covariance in the neighbourhood of the unit root. *Scand. J. Stat. 42* (2015), 234–255.
[18] Miao, Y., and Yang, G. A moderate deviation principle for m–dependent random variables with unbounded m. *Acta Appl. Math. 104* (2008), 191–199.
[19] Phillips, P. C. B., and Lee, J. H. Limit theory for VARs with mixed roots near unity. *Economet. Rev. 34* (2015), 1034–1055.
[20] Phillips, P. C. B., and Magdalinos, T. Limit theory for moderate deviations from a unit root. *J. Econometrics. 136* (2007), 115–130.
[21] Worms, J. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab. 4* (1999), 1–28.
[22] Wu, W. B., and Zhao, Z. Moderate deviations for stationary processes. *Stat. Sinica. 18* (2008), 769–782.

Laboratoire angevin de recherche en mathématiques, LAREMA, UMR 6093, CNRS, UNIV Angers, SFR MathSTIC, 2 Bd Lavoisier, 49045 Angers Cedex 01, France.
Email address: frederic.proia@univ-angers.fr

1.2 Processus à coefficients aléatoires

Dans cette section, nous passons en revue l'article Proïa et Soltane (2018), publié dans *Mathematical Methods of Statistics* et dans lequel nous développons un modèle autorégressif d'ordre $p = 1$ à coefficients aléatoires et autocorrélés, afin de montrer que l'estimation par moindres carrés qui en découle n'est plus consistante et qu'elle doit être corrigée. Nous résumons également sans en fournir l'intégralité l'article Proïa et Soltane (2021), publié dans *Statistics & Probability Letters* et qui vise à généraliser les conclusions précédentes à $p \geqslant 1$.

Résumé

Considérons une trajectoire $(X_0, \ldots, X_n)$ issue d'un modèle autorégressif de la forme

$$\forall\, n \geqslant 1, \quad X_n = \theta_n\, X_{n-1} + \varepsilon_n \tag{1.6}$$

où la suite $(\theta_n)_{n\geqslant 1}$ est formée de coefficients aléatoires et $(\varepsilon_n)_{n\geqslant 1}$ est un bruit blanc de variance $\sigma^2 > 0$. Traditionnellement on exprime les coefficients sous la forme $\theta_n = \theta + \eta_n$ avec $(\eta_n)_{n\geqslant 1}$ un autre bruit blanc indépendant du précédent et de variance $\tau^2 \geqslant 0$, de sorte que $\mathbb{E}[\theta_n] = \theta$ pour tout n et que la présence d'aléa dans les coefficients se ramène au test de $\mathcal{H}_0$: "$\tau^2 = 0$" contre $\mathcal{H}_1$: "$\tau^2 > 0$". On sait depuis longtemps, voir par exemple Nicholls et Quinn (1981), qu'en régime stationnaire (dont la condition s'écrit ici $\theta^2 + \tau^2 < 1$), l'estimation par moindres carrés est fortement consistante pour θ, l'espérance des coefficients, à condition que le processus admette des moments d'ordre 2, et qu'elle est asymptotiquement normale lorsque ce dernier admet des moments d'ordre 4. Nous avons souhaité dans cette étude remettre en question l'hypothèse selon laquelle les coefficients forment eux-mêmes un bruit blanc (centré en θ), hypothèse qui, dans un cadre chronologique, paraît peu crédible. Reprenons donc le modèle (1.6) mais en considérant la suite $(\theta_n)_{n\geqslant 1}$ comme stationnaire et présentant une autocorrélation d'ordre 1. Selon (Brockwell et Davis, 2006, Prop. 3.2.1), cela est équivalent à l'existence d'un bruit blanc $(\eta_n)_{n\geqslant 1}$ tel que les coefficients sont engendrés par une relation MA(1) de la forme

$$\forall\, n \geqslant 1, \quad \theta_n = \theta + \alpha\, \eta_{n-1} + \eta_n \tag{1.7}$$

de sorte que $\mathbb{C}\mathrm{ov}(\theta_n, \theta_{n-1}) = \alpha\, \tau^2$. Ce modèle simple nous permet d'avoir un premier aperçu de l'influence que peut avoir la présence de corrélation dans les coefficients. Sous quelques hypothèses techniques, nous établissons les conditions d'existence d'une solution causale stationnaire à cette équation de récurrence, des moments d'ordre 2 et 4 d'un tel processus et nous en déduisons le comportement asymptotique de l'estimateur usuel des moindres carrés qui, en particulier, perd sa propriété de consistance. Comme corollaire, un test de $\mathcal{H}_0$: "$\alpha = 0$" contre $\mathcal{H}_1$: "$\alpha \neq 0$" permettant de mettre en évidence la présence de corrélation dans les coefficients est proposé. Le contenu de ce travail est donné en intégralité en fin de section. Par ailleurs, précisons sans entrer dans les détails que nous avons poursuivi ces recherches dans le cadre plus général de l'AR(p), voir Proïa et Soltane (2021), c'est-à-dire dans un modèle de la forme

$$\forall\, n \geqslant 1, \quad \Phi_n = (C_\theta + N_{n-1} D_\alpha + N_n)\, \Phi_{n-1} + E_n \tag{1.8}$$

où $E_n = (\varepsilon_n, 0, \ldots, 0)^T$ est un bruit blanc fort p-vectoriel, $\Phi_n = (X_n, \ldots, X_{n-p+1})^T$ avec Φ_0 pour valeur initiale et où le processus admet C_θ comme matrice compagnon (telle que définie dans l'introduction) perturbée par $N_{n-1}D_\alpha + N_n$ contenant des éléments de la forme $\alpha_i\,\eta_{i,\,n-1} + \eta_{i,\,n}$ sur sa première ligne, pour $1 \leqslant i \leqslant p$. En somme, chaque composante de θ_n est perturbée par un MA(1). Nos conclusions sont essentiellement les mêmes dans ce contexte, à savoir que, sous des hypothèses adéquates, l'estimateur des moindres carrés n'est pas consistant mais qu'il converge vers une valeur limite satisfaisant $\theta^* = \theta + \delta$ avec $\alpha = 0 \Rightarrow \delta = 0$ et qu'il reste asymptotiquement normal autour de θ^*. Plus problématique, on montre aussi qu'en général $\theta_i = 0 \not\Rightarrow \theta_i^* = 0$ $(1 \leqslant i \leqslant p)$ et précisément, pour mettre en évidence l'influence néfaste de la présence d'autocorrélation dans les coefficients, un exemple est spécifiquement construit et illustré en simulations dans lequel on pose $\theta_{2,\,n} = 0$ p.s. On explique et on observe que le test de significativité de $\mathcal{H}_0$: "$\theta_2 = 0$" contre $\mathcal{H}_1$: "$\theta_2 \neq 0$" doit être majoritairement rejeté en raison d'une autocorrélation dans $\theta_{1,\,n}$ induisant $\theta_2^* \neq 0$ (voir par exemple la Figure 1.1 issue de Proïa et Soltane (2021)). Dans ce cas, la relation détectée entre X_n et X_{n-2} pourrait être qualifiée de fallacieuse. Les outils techniques utilisés dans ces travaux sont aussi issus de la théorie asymptotique des martingales.

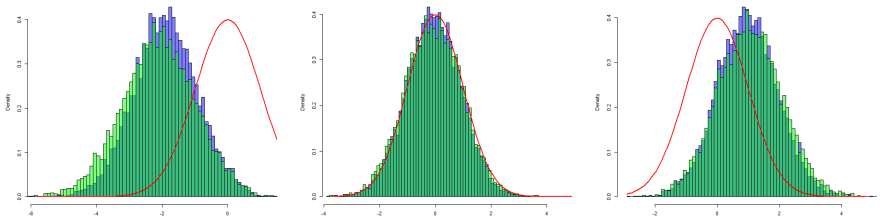


FIGURE 1.1 – Distribution empirique de deux statistiques de test de $\mathcal{H}_0$: "$\theta_2 = 0$" ($N = 10000$ répétitions de taille $n = 500$), avec coefficients aléatoires (bleu) et sans coefficients aléatoires (vert). Ici, $\theta_1 = 0.3$, $\theta_2 = \alpha_2 = \tau_{2,\,2} = 0$, et $\alpha_1 \in \{-0.5, 0, 0.3\}$ de gauche à droite. La courbe en rouge est la densité théorique $\mathcal{N}(0, 1)$.

Perspectives

Lorsque $p = 1$, certaines facilités nous donnent accès à un estimateur corrigé, fortement consistant et asymptotiquement normal pour le couple (θ, α), c'est en particulier ce qui nous permet d'en déduire un test de corrélation dans les coefficients aléatoires. Pour $p > 1$, on peut voir que ce problème n'est pas encore résolu en raison de la complexité calculatoire, sans même envisager des structures de dépendance plus évoluées que les mémoires finies. Si l'on connaît le comportement asymptotique de l'estimation, en revanche on ne sait pas encore estimer θ et α de manière consistante, étape qui pourtant paraît nécessaire pour aboutir à un test similaire au précédent. Par ailleurs ces études ne sont valables qu'en cas de stationnarité, mais par analogie avec la section précédente qui traitait des processus quasi-instables, il pourrait être intéressant de suivre la piste d'un modèle à coefficients aléatoires dont la distribution place la dynamique sur la frontière de la non-stationnarité ou dans son voisinage, afin de mettre en lumière ce que deviennent les problématiques de racines unitaires lorsque les coefficients sont aléatoires sans pour autant être des bruits blancs.

A TEST OF CORRELATION IN THE RANDOM COEFFICIENTS OF AN AUTOREGRESSIVE PROCESS

FRÉDÉRIC PROÏA AND MARIUS SOLTANE

Abstract. A random coefficient autoregressive process is deeply investigated in which the coefficients are correlated. First we look at the existence of a strictly stationary causal solution, we give the second-order stationarity conditions and the autocorrelation function of the process. Then we study some asymptotic properties of the empirical mean and the usual estimators of the process, such as convergence, asymptotic normality and rates of convergence, supplied with the appropriate assumptions on the driving perturbations. Our objective is to get an overview of the influence of correlated coefficients in the estimation step, through a simple model. In particular, the lack of consistency is shown for the estimation of the autoregressive parameter when the independence hypothesis is violated in the random coefficients. Finally, a consistent estimation is given together with a testing procedure for the existence of correlation in the coefficients. While convergence properties rely on the ergodicity, we use a martingale approach to reach most of the results.

Notations and conventions. In the whole paper, I_p is the identity matrix of order p, $[v]_i$ refers to the i–th element of any vector v and M_i to the i–th column of any matrix M. In addition, $\rho(M)$ is the spectral radius of any square matrix M, $M \circ N$ is the Hadamard product between matrices M and N, and $\ln^+ x = \max(\ln x, 0)$. We make the conventions $\sum_{\varnothing} = 0$ and $\prod_{\varnothing} = 1$. Symbols $o(\cdot)$ and $O(\cdot)$ with regard to random sequences will be repeatedly used in the same way as applied to real-valued functions: as $n \to +\infty$, for some positive deterministic rate (v_n), $X_n = o(v_n)$ a.s. means that X_n/v_n converges almost surely to 0 whereas $X_n = O(v_n)$ a.s. means, in the terminology of [9], that for almost all ω, $X_n(\omega) = O(v_n)$, that is $|X_n(\omega)| \leq C(\omega)\, v_n$ for some finite $C(\omega) \geq 0$ and $n \geq N(\omega)$.

1. Introduction and Motivations

In the econometric field, nonlinear time series are now very popular. Our interest lies in some kind of generalization of the standard first-order autoregressive process through random coefficients. The well-known random coefficient autoregressive process RCAR(1) is defined for $t \in \mathbb{Z}$ by

$$X_t = (\theta + \eta_t)X_{t-1} + \varepsilon_t$$

where (ε_t) and (η_t) are uncorrelated white noises. Since the seminal works of Anděl [1] and Nicholls and Quinn [16], stationarity conditions for such processes have been

Key words and phrases. RCAR process, MA process, Random coefficients, Least squares estimation, Stationarity, Ergodicity, Asymptotic normality, Autocorrelation.

widely studied under various assumptions on the moments of (ε_t) and (η_t). Namely, the process was proven to be second-order stationary if $\theta^2 + \tau_2 < 1$ where τ_2 stands for the variance of (η_t). Quite recently, Aue *et al.* [3] have given necessary and sufficient conditions for the existence and uniqueness of a strictly stationary solution of the RCAR(1) process, derived from the more general paper of Brandt [6], and some of our technical assumptions are inspired by their works. However, the flexibility induced by RCAR processes is balanced by the absence of correlation between two consecutive values of the random coefficient. In a time series context, this seems somehow counterintuitive and difficult to argue. Our main objective is precisely to show that the violation of the independence hypothesis in the coefficients, though quite likely for a stochastic phenomenon, leads to a falsification of the whole estimation procedures, and therefore of statistical interpretations. That is the reason why we suggest in this paper an example of random coefficients having a short (finite) memory, in the form of a moving-average dynamic, for which the estimation of the mean value shall be conducted as if they were uncorrelated. For all $t \in \mathbb{Z}$, we consider the first-order autoregressive process given by

$$X_t = \theta_t \, X_{t-1} + \varepsilon_t \tag{1.1}$$

where θ_t is a random coefficient generated by the moving-average structure

$$\theta_t = \theta + \alpha \, \eta_{t-1} + \eta_t. \tag{1.2}$$

This choice of dependence pattern in the coefficients is motivated by Prop. 3.2.1 of [7] which states that any stationary process having finite memory is solution of a moving-average structure. In other words, there exists a white noise such that the random coefficients admit the decomposition given above, and this justifies our interest in (1.2). We can find the foundations of a similar model in Koubkovà [14] or in a far more general way in Brandt [6], but as we will see throughout the paper our objectives clearly diverge. While their works concentrate on the properties of the stationary solution, a large part of this paper focuses on inference. The set of hypotheses that we retain is presented at the end of this introduction, and Section 2 is devoted to the existence, the uniqueness and the stationarity conditions of (X_t). This preliminary study enables us to derive the autocorrelation function of the process. In Section 3, the empirical mean of the process and the usual estimators of θ and σ_2 are investigated, where σ_2 stands for the variance of (ε_t). In particular, we establish some almost sure convergences, asymptotic normalities and rates of convergence, and we also need some results on the fourth-order moments of the process that we deeply examinate. The surprising corollary of these calculations is that the estimation is not consistent for θ as soon as $\alpha \neq 0$, whereas it is well-known that consistency is preserved in the RCAR(1) process. That leads us in Section 4 to build a consistent estimation together with its asymptotic normality, and to derive a statistical procedure for the existence of correlation in the coefficients. In Section 5, we finally prove our results. The estimation of RCAR processes has also been widely addressed in the stationary case, for example by Nicholls and Quinn [15] and later by Schick [19], using either least squares or quasi-maximum likelihood. The crucial point in these works is the strong consistency of the estimation, whereas it appears

in our results that the introduction of correlation in the coefficients is only possible at the cost of consistency. In a general way, our objective is to get an overview of the influence of correlated coefficients in the estimation step through a simple model, to open up new perspectives for more complex structures of dependence. Throughout the paper, we will recall the well-known results related to the first-order stationary RCAR process that are supposed to match with ours for $\alpha = 0$. The reader may find a whole survey in Nicholls and Quinn [17] and without completeness, we also mention the investigations of [18], [13], [11], [12], [4] about inference on RCAR processes, or the unified procedure of Aue and Horváth [2] and references inside. For all $a > 0$, we note the moments

$$\sigma_a = \mathbb{E}[\varepsilon_0^a] \quad \text{and} \quad \tau_a = \mathbb{E}[\eta_0^a].$$

To simplify the calculations, we consider the family of vectors given by

$$U_0 = \begin{pmatrix} 1 \\ 0 \\ \tau_2 \end{pmatrix}, \quad U_1 = \begin{pmatrix} 0 \\ \tau_2 \\ 0 \end{pmatrix}, \quad U_2 = \begin{pmatrix} \tau_2 \\ 0 \\ \tau_4 \end{pmatrix}. \tag{1.3}$$

A particular 3×3 matrix is used all along the study to characterize the second-order properties of the process, it is based on $\{U_0, U_1, U_2\}$ in such a way that

$$M = \begin{pmatrix} \theta^2 + \tau_2 & 2\,\alpha\,\theta & \alpha^2 \\ 2\,\theta\,\tau_2 & 2\,\alpha\,\tau_2 & 0 \\ \theta^2\,\tau_2 + \tau_4 & 2\,\alpha\,\theta\,\tau_2 & \alpha^2\,\tau_2 \end{pmatrix} \quad \text{with} \quad \begin{cases} M_1 = \theta^2\,U_0 + 2\,\theta\,U_1 + U_2 \\ M_2 = 2\,\alpha\,(\theta\,U_0 + U_1) \\ M_3 = \alpha^2\,U_0. \end{cases} \tag{1.4}$$

Similarly, the fourth-order properties of the process rest upon the family of vectors $\{V_0, \ldots, V_4\}$ where

$$V_0 = \begin{pmatrix} 1 \\ 0 \\ \tau_2 \\ 0 \\ \tau_4 \end{pmatrix}, \quad V_1 = \begin{pmatrix} 0 \\ \tau_2 \\ 0 \\ \tau_4 \\ 0 \end{pmatrix}, \quad V_2 = \begin{pmatrix} \tau_2 \\ 0 \\ \tau_4 \\ 0 \\ \tau_6 \end{pmatrix}, \quad V_3 = \begin{pmatrix} 0 \\ \tau_4 \\ 0 \\ \tau_6 \\ 0 \end{pmatrix}, \quad V_4 = \begin{pmatrix} \tau_4 \\ 0 \\ \tau_6 \\ 0 \\ \tau_8 \end{pmatrix}. \tag{1.5}$$

There are used to build the 5×5 matrix H whose columns are defined as

$$\begin{cases} H_1 = \theta^4\,V_0 + 4\,\theta^3\,V_1 + 6\,\theta^2\,V_2 + 4\,\theta\,V_3 + V_4 \\ H_2 = 4\,\alpha\,(\theta^3\,V_0 + 3\,\theta^2\,V_1 + 3\,\theta\,V_2 + V_3) \\ H_3 = 6\,\alpha^2\,(\theta^2\,V_0 + 2\,\theta\,V_1 + V_2) \\ H_4 = 4\,\alpha^3\,(\theta\,V_0 + V_1) \\ H_5 = \alpha^4\,V_0. \end{cases} \tag{1.6}$$

Explicitly,

$$H = \begin{pmatrix} \theta^4 + 6\,\theta^2\,\tau_2 + \tau_4 & 4\,\alpha\,(\theta^3 + 3\,\theta\,\tau_2) & 6\,\alpha^2\,(\theta^2 + \tau_2) & 4\,\alpha^3\,\theta & \alpha^4 \\ 4\,\theta^3\,\tau_2 + 4\,\theta\,\tau_4 & 4\,\alpha\,(3\,\theta^2\,\tau_2 + \tau_4) & 12\,\alpha^2\,\theta\,\tau_2 & 4\,\alpha^3\,\tau_2 & 0 \\ \theta^4\,\tau_2 + 6\,\theta^2\,\tau_4 + \tau_6 & 4\,\alpha\,(\theta^3\,\tau_2 + 3\,\theta\,\tau_4) & 6\,\alpha^2\,(\theta^2\,\tau_2 + \tau_4) & 4\,\alpha^3\,\theta\,\tau_2 & \alpha^4\,\tau_2 \\ 4\,\theta^3\,\tau_4 + 4\,\theta\,\tau_6 & 4\,\alpha\,(3\,\theta^2\,\tau_4 + \tau_6) & 12\,\alpha^2\,\theta\,\tau_4 & 4\,\alpha^3\,\tau_4 & 0 \\ \theta^4\,\tau_4 + 6\,\theta^2\,\tau_6 + \tau_8 & 4\,\alpha\,(\theta^3\,\tau_4 + 3\,\theta\,\tau_6) & 6\,\alpha^2\,(\theta^2\,\tau_4 + \tau_6) & 4\,\alpha^3\,\theta\,\tau_4 & \alpha^4\,\tau_4 \end{pmatrix}.$$

Various hypotheses on the parameters will be required (not always simultaneously) throughout the study, closely related to the distribution of the perturbations.

- (H$_1$) The processes (ε_t) and (η_t) are mutually independent strong white noises such that $\mathbb{E}[\ln^+|\varepsilon_0|] < \infty$ and $\mathbb{E}[\ln|\theta + \alpha\,\eta_0 + \eta_1|] < 0$.
- (H$_2$) $\sigma_{2k+1} = \tau_{2k+1} = 0$ for any $k \in \mathbb{N}$ such that the moments exist.
- (H$_3$) $\sigma_2 > 0$, $\tau_2 > 0$, $\sigma_2 < \infty$, $\tau_4 < \infty$ and $\rho(M) < 1$.
- (H$_4$) $\sigma_4 < \infty$, $\tau_8 < \infty$ and $\rho(H) < 1$.
- (H$_5$) There exists continuous mappings g and h such that $\sigma_4 = g(\sigma_2)$ and $\tau_4 = h(\tau_2)$.

Remark 1.1. *Clearly, (H$_2$) can be replaced by the far less restrictive natural condition $\sigma_1 = \tau_1 = 0$. Considering that all existing odd moments of (ε_t) and (η_t) are zero is only a matter of simplification of the calculations, that are already quite tricky to conduct. An even more general (and possible) study must include the contributions of σ_3, τ_3, τ_5 and τ_7 in the whole calculations.*

Remark 1.2. *(H$_5$) is satisfied in the centered Gaussian case with $g(t) = h(t) = 3\,t^2$. It is also satisfied for most of the distributions used to drive the noise of regression models (centered uniform, Student, Laplace, etc.). Nevertheless, it is a strong assumption only used at the end of the study.*

Short explanations of the remarks appearing in Sections 2 and 3 are given at the beginning of Section 5.

2. Stationarity and Autocorrelation

It is well-known and easy to establish that the sequence of coefficients (θ_t) given by (1.2) is a strictly stationary and ergodic process with mean θ and autocovariance function given by

$$\gamma_\theta(0) = \tau_2\,(1 + \alpha^2), \quad \gamma_\theta(1) = \alpha\,\tau_2 \quad \text{and} \quad \gamma_\theta(h) = 0 \quad (|h| > 1).$$

Clearly, any solution of (1.1) satisfies a recurrence equation, and the first result to investigate is related to the existence of a causal, strictly stationary and ergodic solution.

Theorem 2.1. *Assume that (H$_1$) holds. Then almost surely, for all $t \in \mathbb{Z}$,*

$$X_t = \varepsilon_t + \sum_{k=1}^{\infty} \varepsilon_{t-k} \prod_{\ell=0}^{k-1} (\theta + \alpha\,\eta_{t-\ell-1} + \eta_{t-\ell}). \tag{2.1}$$

In addition, (X_t) is strictly stationary and ergodic.

Proof. See Section 5.2. □

By extension, the same kind of conclusions may be obtained on any process $(\varepsilon_t^a\,\eta_t^b\,X_t^c)$ for $a, b, c \geq 0$, assuming suitable conditions of moments. As a corollary, it will be sufficient to work on $\mathbb{E}[\varepsilon_t^a\,\eta_t^b\,X_t^c]$ in order to identify the asymptotic behavior (for $n \to \infty$) of empirical moments like

$$\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^a\,\eta_t^b\,X_t^c.$$

According to the causal representation of the above theorem, the process is adapted to the filtration defined as

$$\mathcal{F}_t = \sigma((\varepsilon_s, \eta_s),\ s \leq t). \tag{2.2}$$

We are now interested in the existence of the second-order properties of the process, under some additional hypotheses. We derive below its autocorrelation function using the previous notations and letting

$$N = \begin{pmatrix} \theta & \alpha & 0 \\ \tau_2 & 0 & 0 \\ \theta\,\tau_2 & \alpha\,\tau_2 & 0 \end{pmatrix} \quad \text{with} \quad \begin{cases} N_1 = \theta\,U_0 + U_1 \\ N_2 = \alpha\,U_0 \\ N_3 = 0, \end{cases} \tag{2.3}$$

and we take advantage of the calculations to guarantee the unicity of the second-order stationary solution.

Theorem 2.2. *Assume that (H_1)–(H_3) hold. Then, (X_t) is a strictly and second-order stationary process with mean zero and autocovariance function given by*

$$\gamma_X(h) = \sigma_2 \left[N^{|\,h\,|}\,(I_3 - M)^{-1}\,U_0\right]_1 \tag{2.4}$$

for $h \in \mathbb{Z}$. Its autocorrelation function is defined as

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}. \tag{2.5}$$

In addition, this is the unique causal ergodic strictly and second-order stationary solution.

Proof. See Section 5.3. □

Remark 2.1. *Suppose that the process is stationary with second-order moments such that the parameters satisfy $2\,\alpha\,\tau_2 = 1$. Then, (2.4) leads to $\gamma_X(0) = 0$, meaning that (X_t) is a deterministic process. This case is naturally excluded from the study, just like $\sigma_2 = 0$ leading to the same conclusion.*

Remark 2.2. *For $\alpha = 0$, the set of eigenvalues of M is $\{\theta^2 + \tau_2, 0, 0\}$. Thus, the assumption $\rho(M) < 1$ reduces to $\theta^2 + \tau_2 < 1$, which is a well-known result for the stationarity of RCAR(1) processes.*

3. Empirical mean and Usual estimation

Assume that a time series (X_t) generated by (1.1)–(1.2) is observable on the interval $t \in \{0, \ldots, n\}$, for $n \geq 1$. We additionally suppose that X_0 has the strictly stationary and ergodic distribution of the process.

Remark 3.1. *Making the assumption that X_0 has the strictly stationary and ergodic distribution of the process is only a matter of simplification of the calculations. To be complete, assume that (Y_t) is generated by the same recurrence with initial value Y_0. Then for all $t \geq 1$,*

$$X_t - Y_t = (X_0 - Y_0) \prod_{\ell=1}^{t} (\theta + \alpha\,\eta_{\ell-1} + \eta_\ell).$$

For a sufficiently large t and letting $\kappa = \mathbb{E}[\ln|\theta + \alpha\,\eta_0 + \eta_1|] < 0$, it can be shown (see Section 5.1 for details) that, almost surely,

$$|X_t - Y_t| \leq |X_0 - Y_0|\,\mathrm{e}^{\frac{\kappa\,t}{2}}.$$

Then Y_0 could by any random variable satisfying $|X_0 - Y_0| < \infty$ a.s. and having at least as many moments as X_0.

Denote the sample mean by

$$\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{3.1}$$

Then, we have the following result, where the asymptotic variance κ^2 will be explicitly given in (5.18).

Theorem 3.1. *Assume that (H_1)–(H_2) hold. Then as n tends to infinity, we have the almost sure convergence*

$$\bar{X}_n \xrightarrow{\text{a.s.}} 0. \tag{3.2}$$

In addition, if (H_3) also holds, we have the asymptotic normality

$$\sqrt{n}\,\bar{X}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \kappa^2). \tag{3.3}$$

Proof. See Section 5.4. □

Remark 3.2. *For $\alpha = 0$, our calculations lead to*

$$\kappa_0^2 = \frac{\sigma_2\,(1-\theta^2)}{(1-\theta)^2(1-\theta^2-\tau_2)}. \tag{3.4}$$

If in addition $\tau_2 = 0$, we find that

$$\kappa_{00}^2 = \frac{\sigma_2}{(1-\theta)^2} \tag{3.5}$$

which is a result that can be deduced from Thm. 7.1.2 of [7].

Now, consider the estimator given by

$$\widehat{\theta}_n = \frac{\sum_{t=1}^{n} X_{t-1}X_t}{\sum_{t=1}^{n} X_{t-1}^2}. \tag{3.6}$$

It is essential to be well aware that $\widehat{\theta}_n$ is *not* the OLS estimate of θ as soon as $\alpha \neq 0$. This choice of estimate is a consequence of our objectives : to show that an OLS estimation of θ in a standard RCAR(1) model may lead to inappropriate conclusions (due to correlation in the coefficients). Indeed, we shall see in this section that it is not consistent for $\alpha \neq 0$, and we will provide its limiting value. We will also establish that it remains asymptotically normal. This estimator will be described as the *usual* one afterwards. Denote by

$$\theta^* = \frac{\theta}{1 - 2\,\alpha\,\tau_2} \tag{3.7}$$

and recall that $2\,\alpha\,\tau_2 \neq 1$. The asymptotic variance ω^2 in the central limit theorem will be built step by step in Section 5.5 and given in (5.39).

Theorem 3.2. *Assume that (H_1)–(H_3) hold. Then as n tends to infinity, we have the almost sure convergence*

$$\widehat{\theta}_n \xrightarrow{\text{a.s.}} \theta^*. \tag{3.8}$$

In addition, if (H_4) holds, we have the asymptotic normality

$$\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \omega^2). \tag{3.9}$$

Proof. See Section 5.5. □

Remark 3.3. *For $\alpha = 0$, $\theta^* = \theta$ and, as it is well-known, the estimation is consistent for θ. In addition, the coefficients matrix K defined in* (A.4) *takes the very simplified form where each term is zero except $K_{11} = \sigma_2$ and $K_{22} = \tau_2$. Similarly, only the first columns of M and H are nonzero. Then, letting $\lambda_0 = \mathbb{E}[X_t^2] = \gamma_X(0)$ and $\delta_0 = \mathbb{E}[X_t^4]$ as in the associated proof, the asymptotic variance is now*

$$\omega_0^2 = \frac{\sigma_2}{\lambda_0} + \frac{\tau_2\,\delta_0}{\lambda_0^2}.$$

One can check that this is a result of Thm. 4.1 in [15]*, in the particular case of the RCAR(1) process but under more natural hypotheses (they assume that $\mathbb{E}[X_t^4] < \infty$ while we derive it from some moments conditions on the noises). Explicitly, it is given by*

$$\omega_0^2 = \frac{(1-\theta^2-\tau_2)\,(\tau_2\,\sigma_4\,(\theta^2+\tau_2-1) + \sigma_2^2\,(\theta^4+\tau_4-6\,\tau_2^2-1))}{\sigma_2^2\,(\theta^4+\tau_4+6\,\theta^2\,\tau_2-1)}. \tag{3.10}$$

If in addition $\tau_2 = \tau_4 = 0$, we find that

$$\omega_{00}^2 = 1 - \theta^2 \tag{3.11}$$

which is a result stated in Prop. 8.10.1 of [7]*, for example.*

Remark 3.4. *For $\alpha = 0$, the set of eigenvalues of H is $\{\theta^4 + 6\,\theta^2\,\tau_2 + \tau_4, 0, 0, 0, 0\}$. Thus, the assumption $\rho(H) < 1$ reduces to $\theta^4 + 6\,\theta^2\,\tau_2 + \tau_4 < 1$, which may be seen as a condition of existence of fourth-order moments for the RCAR(1) process.*

Theorem 3.3. *Assume that (H_1)–(H_4) hold. Then as n tends to infinity, we have the rates of convergence*

$$\frac{1}{\ln n}\sum_{t=1}^{n}\left(\widehat{\theta}_t - \theta^*\right)^2 \xrightarrow{\text{a.s.}} \omega^2 \tag{3.12}$$

and

$$\limsup_{n\to+\infty}\ \frac{n}{2\,\ln\ln n}\left(\widehat{\theta}_n - \theta^*\right)^2 = \omega^2 \quad \text{a.s.} \tag{3.13}$$

Proof. See Section 5.6. □

Remark 3.5. *The above theorem leads to the usual rate of convergence for the estimation of parameters driving stable models,*

$$\left(\widehat{\theta}_n - \theta^*\right)^2 = O\left(\frac{\ln\ln n}{n}\right) \quad \text{a.s.} \tag{3.14}$$

Remark 3.6. *Even if it is of reduced statistical interest, the same rates of convergence may be reached for* $\bar{X}_n$.

Finally we build the residual set given, for all $1 \leq t \leq n$, by

$$\widehat{\varepsilon}_t = X_t - \widehat{\theta}_n \, X_{t-1}. \tag{3.15}$$

The usual estimator of σ_2 is defined as

$$\widehat{\sigma}_{2,n} = \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_t^{\,2}. \tag{3.16}$$

Denote by

$$\sigma_2^* = \left(1 - (\theta^*)^2\right) \gamma_X(0). \tag{3.17}$$

Theorem 3.4. *Assume that* (H_1)*–*(H_3) *hold. Then as* n *tends to infinity, we have the almost sure convergence*

$$\widehat{\sigma}_{2,n} \xrightarrow{\text{a.s.}} \sigma_2^*. \tag{3.18}$$

Proof. By ergodicity, the development of $\widehat{\sigma}_{2,n}$ in (3.16) leads to

$$\widehat{\sigma}_{2,n} \xrightarrow{\text{a.s.}} \left(1 + (\theta^*)^2\right) \gamma_X(0) - 2\,\theta^* \, \gamma_X(1).$$

But the definition of $\widehat{\theta}_n$ in (3.6) also implies $\gamma_X(1) = \theta^* \, \gamma_X(0)$, leading to σ_2^*. □

Remark 3.7. *For* $\alpha = 0$, (3.17) *becomes*

$$\sigma_{2,0}^* = \frac{\sigma_2 \, (1 - \theta^2)}{1 - \theta^2 - \tau_2}. \tag{3.19}$$

In their work, Nicholls and Quinn [15] *have taken into consideration the fact that this estimator of* σ_2 *was not consistent, that is the reason why they suggested a modified estimator that we will take up in the next section. Now if* $\tau_2 = 0$, *we reach the well-known consistency.*

4. A test for correlation in the coefficients

We now apply a Yule-Walker approach up to the second-order autocorrelation. Using the notations of Theorem 2.2 and letting $\gamma = \alpha \, \tau_2$,

$$\left\{ \begin{array}{rcl} (1 - 2\,\rho_X^2(1))\,\theta & = & (1 - 2\,\rho_X(2))\,\rho_X(1) \\ (1 - 2\,\rho_X^2(1))\,\gamma & = & \rho_X(2) - \rho_X^2(1). \end{array} \right.$$

By ergodicity, a consistent estimation of $\theta^* = \rho_X(1)$ and $\vartheta^* = \rho_X(2)$ is achieved *via*

$$\widehat{\theta}_n = \frac{\sum_{t=1}^{n} X_{t-1} X_t}{\sum_{t=1}^{n} X_{t-1}^2} \quad \text{and} \quad \widehat{\vartheta}_n = \frac{\sum_{t=2}^{n} X_{t-2} X_t}{\sum_{t=2}^{n} X_{t-2}^2} \tag{4.1}$$

respectively. We define the mapping from $[-1\,;\,1] \backslash \{\pm \frac{1}{\sqrt{2}}\} \times [-1\,;\,1]$ to $\mathbb{R}^2$ as

$$f : (x, y) \mapsto \left(\frac{(1 - 2y)x}{1 - 2x^2}, \; \frac{y - x^2}{1 - 2x^2} \right) \tag{4.2}$$

and the new couple of estimates

$$(\widetilde{\theta}_n,\ \widetilde{\gamma}_n) = f(\widehat{\theta}_n, \widehat{\vartheta}_n). \tag{4.3}$$

To be consistent with (4.2), we assume in the sequel that $\sqrt{2}\,\theta \neq \pm(1 - 2\,\alpha\,\tau_2)$. We also assume that $\psi_0^0 \neq 0$, where ψ_0^0 is described below. Since it seems far too complicated, we do not give any reduced form to the latter hypothesis, instead we gather in $\Theta^* = \{\sqrt{2}\,\theta = \pm(1 - 2\,\alpha\,\tau_2)\} \cup \{\psi_0^0 = 0\}$ the pathological cases and we pick the parameters outside Θ^* to conclude our study. It obviously follows that $\widetilde{\theta}_n \xrightarrow{\text{a.s.}} \theta$ and $\widetilde{\gamma}_n \xrightarrow{\text{a.s.}} \gamma$. In the following theorem, we establish the asymptotic normality of these new estimates, useful for the testing procedure. We denote by ∇f the Jacobian matrix of f.

Theorem 4.1. *Assume that (H_1)–(H_4) hold. Then as n tends to infinity, we have the asymptotic normality*

$$\sqrt{n}\begin{pmatrix}\widetilde{\theta}_n - \theta \\ \widetilde{\gamma}_n - \gamma\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Psi) \tag{4.4}$$

where Σ is a covariance given in (5.56) *and*

$$\Psi = \nabla^T f(\theta^*, \vartheta^*)\,\Sigma\,\nabla f(\theta^*, \vartheta^*). \tag{4.5}$$

Proof. See Section 5.7. □

Assuming random coefficients (that is, $\tau_2 > 0$), note that $\gamma = 0 \Leftrightarrow \alpha = 0$. Our last objective is to build a testing procedure for

$$\mathcal{H}_0 : \text{“}\alpha = 0\text{”} \quad \text{vs} \quad \mathcal{H}_1 : \text{“}\alpha \neq 0\text{”}. \tag{4.6}$$

As it is explained in Remark 5.1, despite its complex structure, Ψ only depends on the parameters. Let $\psi = \psi(\theta, \alpha, \{\tau_k\}_{2,4,6,8}, \{\sigma_\ell\}_{2,4})$ be the the lower right element of Ψ, and $\psi^0 = \psi(\theta, 0, \{\tau_k\}_{2,4,6,8}, \{\sigma_\ell\}_{2,4})$. The explicit calculation under $\mathcal{H}_0$ gives $\theta^* = \theta$, $\vartheta^* = \theta^2$ and

$$\psi^0 = \frac{\psi_0^0}{(1 - 2\,\theta^2)^2\,\sigma_2^2\,(\theta^4 + 6\,\theta^2\,\tau_2 + \tau_4 - 1)} \tag{4.7}$$

where the numerator is given by

$$\begin{aligned}\psi_0^0 \;=\; & (\tau_2 + \theta^2 - 1)\,\big[\sigma_4\,\tau_2\,((6\,\theta^2 - 1)\,\tau_2^2 + (8\,\theta^4 - 9\,\theta^2 + 1)\,\tau_2 \\ & + 2\,\theta^2\,(\theta^2 - 1)^2) + \sigma_2^2\,\tau_2\,(-36\,\tau_2^2\,\theta^2 + 6\,\tau_2^2 - 12\,\tau_2\,\theta^4 \\ & + 12\,\tau_2\,\theta^2 - 6\,\theta^6 + 17\,\theta^4 + 6\,\tau_4\,\theta^2 - 12\,\theta^2 - \tau_4 + 1) \\ & + \sigma_2^2\,(\theta^6 - \theta^4 + \theta^2\,\tau_4 - \theta^2 - \tau_4 + 1)\big]\end{aligned}$$

and assumed to be nonzero (by excluding Θ^*). As a corollary, ψ^0 continuously depends on the parameters under our additional hypothesis (see Remark 3.4). Suppose also that (H_5) holds, so that $\psi^0 = \psi^0(\theta, \tau_2, \sigma_2)$, and consider

$$\widehat{\psi}_n^0 = \psi^0(\bar{\theta}_n, \bar{\tau}_{2,n}, \bar{\sigma}_{2,n})$$

where $\bar{\theta}_n$ is either $\widehat{\theta}_n$ or $\widetilde{\theta}_n$, and $(\bar{\tau}_{2,n}, \bar{\sigma}_{2,n})$ is the couple of estimates suggested by [15] in formulas (3.6) and (3.7) respectively, also given in [13]. They are defined as

$$\bar{\tau}_{2,n} = \frac{\sum_{t=1}^{n}(Z_t - \bar{Z}_n)\,\widehat{\varepsilon}_t^{\,2}}{\sum_{t=1}^{n}(Z_t - \bar{Z}_n)^2} \quad \text{and} \quad \bar{\sigma}_{2,n} = \widehat{\sigma}_{2,n} - \bar{Z}_n\,\bar{\tau}_{2,n} \tag{4.8}$$

where $(\widehat{\varepsilon}_t)$ is the residual set built in (3.15), $\widehat{\sigma}_{2,n}$ is given in (3.16) and for $t \in \{1, \ldots, n\}$, $Z_t = X_t^2$. Thm. 4.2 of [15] gives their consistency as soon as the RCAR(1) process has fourth-order moments. Furthermore, our study gives the consistency of $\bar{\theta}_n$ under $\mathcal{H}_0$. We deduce from Slutsky's lemma that

$$\widehat{\psi}_n^0 \xrightarrow{\text{a.s.}} \psi^0 > 0 \quad \text{and} \quad \frac{n\left(\widetilde{\gamma}_n\right)^2}{\widehat{\psi}_n^0} \xrightarrow{\mathcal{D}} \chi_1^2 \tag{4.9}$$

if $\mathcal{H}_0$ is true, where χ_1^2 has a chi-square distribution with one degree of freedom, whereas under $\mathcal{H}_1$ the test statistic diverges (almost surely). The introduction of (H_5) enables to choose

$$\bar{\sigma}_{4,n} = g(\bar{\sigma}_{2,n}) \quad \text{and} \quad \bar{\tau}_{4,n} = h(\bar{\tau}_{2,n})$$

as consistent estimations of the related moments. Comparing the test statistic with the quantiles of χ_1^2 may constitute the basis of a test for the existence of correlation in the random coefficients of an autoregressive process. To conclude, we have shown through this simple model that the introduction of correlation in the coefficients is a significative issue in relation to the inference procedure. And yet, in a time series context it seems quite natural to take account of autocorrelation in the random coefficients, this is an incitement to put statistical conclusions into perspective dealing with estimation and testing procedures of RCAR models. The most challenging extensions for future studies seem to rely on more complex dependency structures in the coefficients, on the consideration of more autoregressions in the model, and of course on the behavior of the process under instability and unit root issues. The testing procedure for correlation in the random coefficients should also be studied on an empirical basis, this is an ongoing investigation.

Acknowledgments. The authors thank the Associate Editor and the anonymous Reviewer for the suggestions and very constructive comments which helped to improve substantially the paper.

5. Proofs of the main results

In this section, we develop the whole proofs of our results. The fundamental tools related to ergodicity may be found in Thm. 3.5.8 of [21] or in Thm. 1.3.3 of [22]. We will repeatedly have to deal with $\mathbb{E}[\eta_t^a(\theta + \eta_t)^b]$ for $a, b \in \{0, \ldots, 4\}$, so we found useful to summarize beforehand the associated values under (H_2) in Table 1 below. For the sake of clarity, we postpone to the appendix the numerous constants that will be used thereafter. We start by giving some short explanations related to the remarks appearing in Sections 2 and 3.

5.1. About the remarks of Sections 2 and 3.

$a\backslash b$	0	1	2	3	4
0	1	θ	$\theta^2+\tau_2$	$\theta^3+3\,\theta\,\tau_2$	$\theta^4+6\,\theta^2\,\tau_2+\tau_4$
1	0	τ_2	$2\,\theta\,\tau_2$	$3\,\theta^2\,\tau_2+\tau_4$	$4\,\theta^3\,\tau_2+4\,\theta\,\tau_4$
2	τ_2	$\theta\,\tau_2$	$\theta^2\,\tau_2+\tau_4$	$\theta^3\,\tau_2+3\,\theta\,\tau_4$	$\theta^4\,\tau_2+6\,\theta^2\,\tau_4+\tau_6$
3	0	τ_4	$2\,\theta\,\tau_4$	$3\,\theta^2\,\tau_4+\tau_6$	$4\,\theta^3\,\tau_4+4\,\theta\,\tau_6$
4	τ_4	$\theta\,\tau_4$	$\theta^2\,\tau_4+\tau_6$	$\theta^3\,\tau_4+3\,\theta\,\tau_6$	$\theta^4\,\tau_4+6\,\theta^2\,\tau_6+\tau_8$

TABLE 1. $\mathbb{E}[\eta_t^a(\theta+\eta_t)^b]$ for $a,b\in\{0,\dots,4\}$.

5.1.1. *Remark 2.1.* Indeed, the explicit calculation of $\gamma_X(0)$ based on (2.4) leads to

$$\gamma_X(0)=\frac{\sigma_2\,(2\,\alpha\,\tau_2-1)}{d(\theta,\alpha,\tau_2,\tau_4)}$$

for some denominator satisfying $d(\theta,\alpha,\tau_2,\tau_4)=2\,\theta^2$ when $2\,\alpha\,\tau_2=1$. It follows that should this assumption be true under second-order stationarity, the process would be deterministic.

5.1.2. *Remark 3.1.* The objective here is to show that the difference between the process starting at X_0 having the strictly stationary and ergodic distribution and the same process starting at some Y_0 is (a.s.) negligible provided very weak assumptions on Y_0. Following the idea of Lem. 1 in [3] and using the ergodic theorem, we obtain that for a sufficiently large t, almost surely

$$\frac{1}{t}\sum_{\ell=1}^{t}\ln|\theta+\alpha\,\eta_{\ell-1}+\eta_\ell|\le\frac{\kappa}{2}<0.$$

Hence, the asymptotic decrease of $\prod_{\ell=1}^{t}|\theta+\alpha\,\eta_{\ell-1}+\eta_\ell|$ is exponentially fast with t under (H_1) and the upper bound of $|X_t-Y_t|\le|X_0-Y_0|\,\mathrm{e}^{\frac{\kappa\,t}{2}}$ enables to retain weak assumptions on Y_0 so that $X_t-Y_t=o(1)$ a.s.

5.1.3. *Remark 3.2.* In the particular case where $\alpha=\tau_2=0$ (that is, in the stable AR(1) process), Thm. 7.1.2 of [7] states that $\sqrt{n}\,\bar{X}_n$ is asymptotically normal with mean 0 and variance given by

$$\sum_{h\,\in\,\mathbb{Z}}\gamma_X(h)=\sigma_2\Big(\sum_{k=0}^{+\infty}\theta^k\Big)^2=\frac{\sigma_2}{(1-\theta)^2}.$$

Thus, κ_{00}^2 implied by our results is coherent from that point of view.

5.1.4. *Remark 3.3.* Like in the previous remark, Prop. 8.10.1 of [7] states that, for $\alpha=\tau_2=0$, the OLS estimator of θ is asymptotically normal with rate $\sqrt{n}$, mean 0 and variance given by $1-\theta^2$, which corresponds to ω_{00}^2. Now if $\tau_2>0$, Thm. 4.1 of [15], and especially formula (4.1), gives the asymptotic variance as a function of $\mathbb{E}[X_t^2]$ and $\mathbb{E}[X_t^4]$ as detailed in Rem. 3.3. Our study enables to identify ω_0^2 as a function of the parameters by injecting $\alpha=0$ into λ_0 and δ_0 that are computed in (5.10) and (5.30), respectively.

5.2. **Proof of Theorem 2.1.** The existence of the almost sure causal representation of (X_t) under (H_1) is a corollary of Thm. 1 of [6]. Indeed, (θ_t) is a stationary and ergodic MA(1) process independent of (ε_t), itself obviously stationary and ergodic. Let us give more details. First, hypotheses (H_1) enable to make use of the same proof as [3] where the ergodic theorem replaces the strong law of large numbers to reach formula (6), and to establish that (2.1) is the limit of a convergent series (with probability 1). Then for all $t \in \mathbb{Z}$,

$$\begin{aligned}\theta_t\, X_{t-1} &= (\theta + \alpha\,\eta_{t-1} + \eta_t)\left[\varepsilon_{t-1} + \sum_{k=1}^{\infty} \varepsilon_{t-k-1} \prod_{\ell=0}^{k-1} (\theta + \alpha\,\eta_{t-\ell-2} + \eta_{t-\ell-1})\right] \\ &= \sum_{k=1}^{\infty} \varepsilon_{t-k} \prod_{\ell=0}^{k-1} (\theta + \alpha\,\eta_{t-\ell-1} + \eta_{t-\ell}) \;=\; X_t - \varepsilon_t\end{aligned}$$

meaning that (2.1) is a solution to the recurrence equation. Finally, the strict stationarity and ergodicity of (X_t) may be obtained following the same reasoning as in [15]. Indeed, the causal representation (2.1) shows that there exists ϕ independent of t such that for all $t \in \mathbb{Z}$,

$$X_t = \phi((\varepsilon_t, \eta_t), (\varepsilon_{t-1}, \eta_{t-1}), \ldots).$$

The set $((\varepsilon_t, \eta_t), (\varepsilon_{t-1}, \eta_{t-1}), \ldots)$ being made of independent and identically distributed random vectors, (X_t) is strictly stationary. The ergodicity follows from Thm. 1.3.3 of [22]. □

5.3. **Proof of Theorem 2.2.** Ergodicity and strict stationarity come from Theorem 2.1. We consider the causal representation (2.1). First, since (ε_t) and (η_t) are uncorrelated white noises, for all $t \in \mathbb{Z}$,

$$\mathbb{E}[X_t] = 0. \tag{5.1}$$

To establish the autocovariance function of (X_t), we have beforehand to establish a technical lemma related to the second-order properties of the process. For all $k, h \in \mathbb{N}^*$, consider the sequence

$$\begin{aligned} u_{0,h}^{(a)} &= \mathbb{E}[\eta_h^a\, \theta_h \ldots \theta_1], \\ u_{k,0}^{(a)} &= \mathbb{E}[\eta_k^a\, \theta_k^2 \ldots \theta_1^2], \\ u_{k,h}^{(a)} &= \mathbb{E}[\eta_{k+h}^a\, \theta_{k+h} \ldots \theta_{k+1}\, \theta_k^2 \ldots \theta_1^2], \end{aligned}$$

where $a \in \{0, 1, 2\}$, and build

$$U_{k,h} = \begin{pmatrix} u_{k,h}^{(0)} \\ u_{k,h}^{(1)} \\ u_{k,h}^{(2)} \end{pmatrix}. \tag{5.2}$$

Thereafter, M, N and U_0 refer to (1.4), (2.3) and (1.3), respectively.

Lemma 5.1. *Assume that (H_1)–(H_3) hold. Then, for all $h, k \in \mathbb{N}$,*

$$U_{k,h} = N^h\, M^k\, U_0 \tag{5.3}$$

with the convention that $U_{0,0} = U_0$.

Proof. In the whole proof, $(\mathcal{F}_t)$ is the filtration defined in (2.2) and Table 1 may be read to compute the coefficients appearing in the calculations. The coefficients θ_{k+h-1}, θ_{k+h-2}, ... are $\mathcal{F}_{k+h-1}$–measurable. Hence for $h \geq 1$,

$$\begin{aligned}
u_{k,h}^{(0)} &= \mathbb{E}[\theta_{k+h-1} \ldots \theta_{k+1}\, \theta_k^2 \ldots \theta_1^2\, \mathbb{E}[\theta_{k+h} \,|\, \mathcal{F}_{k+h-1}]] \\
&= \theta\, u_{k,h-1}^{(0)} + \alpha\, u_{k,h-1}^{(1)}, \\
u_{k,h}^{(1)} &= \mathbb{E}[\theta_{k+h-1} \ldots \theta_{k+1}\, \theta_k^2 \ldots \theta_1^2\, \mathbb{E}[\eta_{k+h}\, \theta_{k+h} \,|\, \mathcal{F}_{k+h-1}]] \\
&= \tau_2\, u_{k,h-1}^{(0)}, \\
u_{k,h}^{(2)} &= \mathbb{E}[\theta_{k+h-1} \ldots \theta_{k+1}\, \theta_k^2 \ldots \theta_1^2\, \mathbb{E}[\eta_{k+h}^2\, \theta_{k+h} \,|\, \mathcal{F}_{k+h-1}]] \\
&= \theta\, \tau_2\, u_{k,h-1}^{(0)} + \alpha\, \tau_2\, u_{k,h-1}^{(1)}.
\end{aligned}$$

We get the matrix formulation $U_{k,h} = N\, U_{k,h-1}$. It follows that, for $h \in \mathbb{N}$,

$$U_{k,h} = N^h\, U_{k,0}. \tag{5.4}$$

The next step is to compute $U_{k,0}$, and we will use the same lines. For $k \geq 1$,

$$\begin{aligned}
u_{k,0}^{(0)} &= \mathbb{E}[\theta_{k-1}^2 \ldots \theta_1^2\, \mathbb{E}[\theta_k^2 \,|\, \mathcal{F}_{k-1}]] \\
&= (\theta^2 + \tau_2)\, u_{k-1,0}^{(0)} + 2\,\alpha\,\theta\, u_{k-1,0}^{(1)} + \alpha^2\, u_{k-1,0}^{(2)}, \\
u_{k,0}^{(1)} &= \mathbb{E}[\theta_{k-1}^2 \ldots \theta_1^2\, \mathbb{E}[\eta_k\, \theta_k^2 \,|\, \mathcal{F}_{k-1}]] \\
&= 2\,\theta\,\tau_2\, u_{k-1,0}^{(0)} + 2\,\alpha\,\tau_2\, u_{k-1,0}^{(1)}, \\
u_{k,0}^{(2)} &= \mathbb{E}[\theta_{k-1}^2 \ldots \theta_1^2\, \mathbb{E}[\eta_k^2\, \theta_k^2 \,|\, \mathcal{F}_{k-1}]] \\
&= (\theta^2\,\tau_2 + \tau_4)\, u_{k-1,0}^{(0)} + 2\,\alpha\,\theta\,\tau_2\, u_{k-1,0}^{(1)} + \alpha^2\,\tau_2\, u_{k-1,0}^{(2)}.
\end{aligned}$$

Thus, (5.4) becomes

$$U_{k,h} = N^h\, M^{k-1}\, U_{1,0}$$

where the initial vector $U_{1,0}$ is given by

$$\begin{aligned}
u_{1,0}^{(0)} &= \mathbb{E}[\theta_1^2] \;=\; (\theta^2 + \tau_2) + \alpha^2\,\tau_2, \\
u_{1,0}^{(1)} &= \mathbb{E}[\eta_1\, \theta_1^2] \;=\; 2\,\theta\,\tau_2, \\
u_{1,0}^{(2)} &= \mathbb{E}[\eta_1^2\, \theta_1^2] \;=\; (\theta^2\,\tau_2 + \tau_4) + \alpha^2\,\tau_2^2.
\end{aligned}$$

It is then not hard to conclude that, for all $k \in \mathbb{N}^*$ and $h \in \mathbb{N}$,

$$U_{k,h} = N^h\, M^k\, U_0.$$

For $k = 0$, a similar calculation based on the initial values $u_{0,h}^{(a)}$ for $a \in \{0, 1, 2\}$ leads to $U_{0,h} = N^h\, U_0$, implying that (5.3) holds for all $k, h \in \mathbb{N}$. □

Corollary 5.1. *Assume that* (H_1)*–*(H_3) *hold. Then, the second-order properties of* (X_t) *are such that, for all* $a \in \{0, 1, 2\}$,

$$\mathbb{E}[\eta_t^a\, X_t^2] < \infty.$$

Proof. For all $t \in \mathbb{Z}$ and $k \geq 1$, denote by

$$\Lambda_t = \begin{pmatrix} 1 \\ \eta_t \\ \eta_t^2 \end{pmatrix} \quad \text{and} \quad P_{t,\,k} = \prod_{i=0}^{k-1} \theta_{t-i} \tag{5.5}$$

with $P_{t,\,0} = 1$. Since (ε_t) and (η_t) are uncorrelated white noises, using the causal representation (2.1) and letting $h = 0$,

$$\mathbb{E}[\Lambda_t\, X_t^2] = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \mathbb{E}[\Lambda_t\, P_{t,\,k}\, P_{t,\,\ell}\, \varepsilon_{t-k}\, \varepsilon_{t-\ell}] = \sigma_2 \sum_{k=0}^{\infty} M^k\, U_0 = \sigma_2\, (I_3 - M)^{-1}\, U_0$$

as a consequence of the strict stationarity of (θ_t). We remind that, under (H_3), it is well-known (see *e.g.* [10]) that $I_3 - M$ is invertible and that

$$\sum_{k=0}^{\infty} M^k = (I_3 - M)^{-1}.$$

□

Let us return to the proof of Theorem 2.2. From Lemma 5.1 and Corollary 5.1, we are now able to evaluate the autocovariance function of (X_t). For $h \in \mathbb{N}$,

$$\mathbb{C}\mathrm{ov}(X_t, X_{t-h}) = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \mathbb{E}[P_{t,\,k}\, P_{t-h,\,\ell}\, \varepsilon_{t-k}\, \varepsilon_{t-h-\ell}].$$

We get

$$\gamma_X(h) = \sigma_2 \sum_{k=0}^{\infty} \mathbb{E}[P_{t,\,k+h}\, P_{t-h,\,k}] = \sigma_2 \left(\mathbb{E}[P_{t,\,h}] + \sum_{k=1}^{\infty} u_{k,h}^{(0)} \right) = \sigma_2 \Big[\sum_{k=0}^{\infty} U_{k,h} \Big]_1.$$

From Lemma 5.1,

$$\gamma_X(h) = \sigma_2 \left[N^h\, (I_3 - M)^{-1}\, U_0 \right]_1.$$

We conclude using the fact that γ_X does not depend on t. For all $t \in \mathbb{Z}$ and $h \in \mathbb{N}$, $\gamma_X(h) = \mathbb{C}\mathrm{ov}(X_{t-h}, X_t) = \mathbb{C}\mathrm{ov}(X_t, X_{t+h})$, which shows that the above reasoning still holds for $h \in \mathbb{Z}$, replacing h by $|h|$. Now suppose that (W_t) is another causal ergodic strictly and second-order stationary solution. There exists φ independent of t such that for all $t \in \mathbb{Z}$,

$$X_t - W_t = \varphi((\varepsilon_t, \eta_t), (\varepsilon_{t-1}, \eta_{t-1}), \ldots)$$

and necessarily, $(X_t - W_t)$ is also a strictly stationary process having second-order moments. Let $e^{(a)} = \mathbb{E}[\eta_t^a\, (X_t - W_t)^2]$, for $a \in \{0, 1, 2\}$. From the same calculations and exploiting the second-order stationarity of $(X_t - W_t)$, it follows that

$$\begin{pmatrix} e^{(0)} \\ e^{(1)} \\ e^{(2)} \end{pmatrix} = M \begin{pmatrix} e^{(0)} \\ e^{(1)} \\ e^{(2)} \end{pmatrix}$$

implying, if $(e^{(0)}\ e^{(1)}\ e^{(2)}) \neq 0$, that 1 is an eigenvalue of M. Clearly, this contradicts $\rho(M) < 1$ which is part of (H_3). Thus, $\mathbb{E}[(X_t - W_t)^2]$ must be zero and $X_t = W_t$ a.s.

□

5.4. **Proof of Theorem 3.1.** The convergence to zero is only the application of the ergodic theorem, since we have seen in (5.1) that $\mathbb{E}[X_t] = 0$. Here, only (H_1) and (H_2) are needed. We make the following notations,

$$\begin{aligned}
\bar{M}_n^{(1)} &= \sum_{t=1}^{n} X_{t-1}\left((1+\alpha\,\theta)\,\eta_t + \alpha\,(\eta_t^2 - \tau_2)\right),\\
\bar{M}_n^{(2)} &= \alpha^2 \sum_{t=1}^{n} \eta_{t-1}\,X_{t-1}\,\eta_t,\\
\bar{M}_n^{(3)} &= \sum_{t=1}^{n} (1+\alpha\,\eta_t)\,\varepsilon_t.
\end{aligned}$$

Consider the filtration $(\mathcal{F}_n^*)$ generated by $\mathcal{F}_0^* = \sigma(X_0, \eta_0)$ and, for $n \geq 1$, by

$$\mathcal{F}_n^* = \sigma(X_0, \eta_0, (\varepsilon_1, \eta_1), \ldots, (\varepsilon_n, \eta_n)) \tag{5.6}$$

and let

$$\bar{\mathcal{M}}_n = \begin{pmatrix} \bar{M}_n^{(1)} \\ \bar{M}_n^{(2)} \\ \bar{M}_n^{(3)} \end{pmatrix}. \tag{5.7}$$

Under our hypotheses, $\bar{\mathcal{M}}_n$ is a locally square-integrable real vector $(\mathcal{F}_n^*)$–martingale. We shall make use of the central limit theorem for vector martingales given *e.g.* by Cor. 2.1.10 of [9]. On the one hand, we have to study the asymptotic behavior of the predictable quadratic variation of $\bar{\mathcal{M}}_n$. For all $n \geq 1$, let

$$\langle \bar{\mathcal{M}} \rangle_n = \sum_{t=1}^{n} \mathbb{E}\big[(\Delta\bar{\mathcal{M}}_t)(\Delta\bar{\mathcal{M}}_t)^T \,|\, \mathcal{F}_{t-1}^*\big], \tag{5.8}$$

with $\Delta\bar{\mathcal{M}}_1 = \bar{\mathcal{M}}_1$. To simplify the calculations, we introduce some more notations. The second-order moments of the process are called

$$\mathbb{E}[\Lambda_t\,X_t^2] = \begin{pmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \Lambda \tag{5.9}$$

where Λ_t is given in (5.5), with $\lambda_0 = \gamma_X(0)$. We use the strict stationarity to establish, following Corollary 5.1 under the additional (H_3) hypothesis, that

$$\Lambda = \sigma_2\,(I_3 - M)^{-1}\,U_0 \tag{5.10}$$

and ergodicity immediately leads to

$$\frac{1}{n}\sum_{t=1}^{n} \Lambda_t\,X_t^2 \;\xrightarrow{\text{a.s.}}\; \Lambda. \tag{5.11}$$

Now, we are going to study the asymptotic behavior of $\langle \bar{\mathcal{M}} \rangle_n / n$. First, under our assumptions,

$$\langle \bar{M}^{(1)}, \bar{M}^{(3)} \rangle_n = \langle \bar{M}^{(2)}, \bar{M}^{(3)} \rangle_n = 0.$$

Since the other calculations are very similar we only detail the first one,

$$\begin{aligned}\langle \bar{M}^{(1)} \rangle_n &= \sum_{t=1}^{n} X_{t-1}^2 \, \mathbb{E}\big[\big((1+\alpha\,\theta)\,\eta_t + \alpha\,(\eta_t^2 - \tau_2)\big)^2\big] \\ &= \big((1+\alpha\,\theta)^2\,\tau_2 + \alpha^2\,(\tau_4 - \tau_2^2)\big) \sum_{t=1}^{n} X_{t-1}^2.\end{aligned}$$

We obtain using $\bar{K}$ in (A.2) that

$$\langle \bar{\mathcal{M}} \rangle_n = \bar{K} \circ \sum_{t=1}^{n} \begin{pmatrix} X_t^2 & \eta_t\,X_t^2 & 0 \\ \eta_t\,X_t^2 & \eta_t^2\,X_t^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \bar{R}_n \tag{5.12}$$

where the Hadamard product $\circ$ is used to lighten the formula, and where the remainder $\bar{R}_n$ is made of isolated terms such that, from (5.11),

$$\frac{\bar{R}_n}{n} \xrightarrow{\text{a.s.}} 0. \tag{5.13}$$

We reach these results by computing $\langle \bar{M}^{(i)}, \bar{M}^{(j)} \rangle_n$ for $i, j \in \{1, 2, 3\}$ just as we have done above for some of them, and then by normalizing each sum, leaving the isolated terms in the remainder. For example,

$$\sum_{t=1}^{n} X_{t-1}^2 = \sum_{t=1}^{n} X_t^2 + (X_0^2 - X_n^2).$$

It is then a direct application of the ergodic theorem that gives the $O(n)$ behavior of the sums (and the $o(n)$ behavior of the isolated terms as a consequence), and that enables to identify, by combining (5.11), (5.12) and (5.13), the limiting value

$$\frac{\langle \bar{\mathcal{M}} \rangle_n}{n} \xrightarrow{\text{a.s.}} \bar{K} \circ \bar{\Gamma} \tag{5.14}$$

where $\bar{\Gamma}$ is given by

$$\bar{\Gamma} = \begin{pmatrix} \lambda_0 & \lambda_1 & 0 \\ \lambda_1 & \lambda_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.15}$$

On the other hand, it is necessary to prove that the Lindeberg's condition is satisfied, namely that for all $\varepsilon > 0$,

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\big[\|\Delta \bar{\mathcal{M}}_t\|^2 \, \mathbb{I}_{\{\|\Delta \bar{\mathcal{M}}_t\| \geq \varepsilon \sqrt{n}\}} \,|\, \mathcal{F}_{t-1}^*\big] \xrightarrow{\mathbb{P}} 0 \tag{5.16}$$

as n tends to infinity. By ergodicity and strict stationarity of the increments $(\Delta \bar{\mathcal{M}}_t)$ under the assumption on X_0, it follows that for any $M > 0$,

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\big[\|\Delta \bar{\mathcal{M}}_t\|^2 \, \mathbb{I}_{\{\|\Delta \bar{\mathcal{M}}_t\| \geq M\}} \,|\, \mathcal{F}_{t-1}^*\big] \xrightarrow{\text{a.s.}} \mathbb{E}\big[\|\Delta \bar{\mathcal{M}}_1\|^2 \, \mathbb{I}_{\{\|\Delta \bar{\mathcal{M}}_1\| \geq M\}}\big].$$

Corollary 5.1 implies that $\mathbb{E}[\|\Delta\bar{\mathcal{M}}_1\|^2] < \infty$ and the right-hand side can be made arbitrarily small, which establishes the Lindeberg's condition. From (5.14) and (5.16), we deduce that

$$\frac{\bar{\mathcal{M}}_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \bar{K} \circ \bar{\Gamma}) \tag{5.17}$$

which is nothing but the central limit theorem for vector martingales, as intended. One can notice that the above reasoning is in fact a vector extension of the main result of [5], related to the central limit theorem for martingales having ergodic and stationary increments. Finally, by a tedious but straightforward calculation, one can obtain that

$$\sqrt{n}\,\bar{X}_n = \frac{\Omega_3^T \bar{\mathcal{M}}_n + \bar{r}_n}{(1 - \theta - \alpha\,\tau_2)\sqrt{n}}$$

where $\Omega_3^T = (1\ 1\ 1)$ and $\bar{r}_n = o(\sqrt{n})$ a.s. from (5.11). It remains to apply Slutsky's lemma to conclude that

$$\sqrt{n}\,\bar{X}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \kappa^2)$$

with

$$\kappa^2 = \frac{\Omega_3^T (\bar{K} \circ \bar{\Gamma})\,\Omega_3}{(1 - \theta - \alpha\,\tau_2)^2} \tag{5.18}$$

using the whole notations above. □

5.5. **Proof of Theorem 3.2.** The almost sure convergence essentially relies on the ergodicity of the process. Theorem 2.2 together with the ergodic theorem directly lead to

$$\widehat{\theta}_n \xrightarrow{\text{a.s.}} \frac{\gamma_X(1)}{\gamma_X(0)} = \frac{\left[N\,(I_3 - M)^{-1}\,U_0\right]_1}{\left[(I_3 - M)^{-1}\,U_0\right]_1}$$

as n tends to infinity, but we are interested in the explicit form of the limiting value. From the combined expressions (1.1)–(1.2), it follows that

$$\sum_{t=1}^{n} X_{t-1}X_t = \theta \sum_{t=1}^{n} X_{t-1}^2 + \alpha \sum_{t=1}^{n} \eta_{t-1}\,X_{t-1}^2 + \sum_{t=1}^{n} X_{t-1}^2\,\eta_t + \sum_{t=1}^{n} X_{t-1}\,\varepsilon_t. \tag{5.19}$$

We also note from Corollary 5.1 that, for all $t \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{E}[\eta_t\,X_t^2] &= \mathbb{E}[\theta_t^2\,X_{t-1}^2\,\eta_t] + \mathbb{E}[\varepsilon_t^2\,\eta_t] + 2\,\mathbb{E}[\theta_t\,X_{t-1}\,\varepsilon_t\,\eta_t] \\ &= 2\,\alpha\,\tau_2\,\mathbb{E}[\eta_{t-1}\,X_{t-1}^2] + 2\,\theta\,\tau_2\,\mathbb{E}[X_{t-1}^2].\end{aligned}$$

Thus, by stationarity and ergodicity,

$$\frac{1}{n}\sum_{t=1}^{n} \eta_{t-1}\,X_{t-1}^2 \xrightarrow{\text{a.s.}} \frac{2\,\theta\,\tau_2\,\gamma_X(0)}{1 - 2\,\alpha\,\tau_2}. \tag{5.20}$$

Similarly, $\mathbb{E}[X_{t-1}^2\,\eta_t] = \mathbb{E}[X_{t-1}\,\varepsilon_t] = 0$ and from the ergodic theorem,

$$\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \xrightarrow{\text{a.s.}} \gamma_X(0), \quad \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2\,\eta_t \xrightarrow{\text{a.s.}} 0, \quad \frac{1}{n}\sum_{t=1}^{n} X_{t-1}\,\varepsilon_t \xrightarrow{\text{a.s.}} 0. \tag{5.21}$$

The expression of $\widehat{\theta}_n$ in (3.6) combined with the decomposition (5.19) and the convergences (5.20) and (5.21) give

$$\widehat{\theta}_n \xrightarrow{\text{a.s.}} \theta + \frac{2\,\alpha\,\theta\,\tau_2}{1-2\,\alpha\,\tau_2} = \frac{\theta}{1-2\,\alpha\,\tau_2}.$$

Let us now establish the asymptotic normality. First, we have to study the fourth-order properties of (X_t) and some other technical lemmas are needed. For all $k \in \mathbb{N}^*$, consider the sequences

$$v_k^{(a)} \;=\; \mathbb{E}[\eta_k^a\,\theta_k^4\,\ldots\,\theta_1^4]$$

where $a \in \{0,\ldots,4\}$, and build

$$V_k = \begin{pmatrix} v_k^{(0)} \\ \vdots \\ v_k^{(4)} \end{pmatrix}. \tag{5.22}$$

For the following calculations, H is defined in (1.6) and $\{V_0,\ldots,V_4\}$ in (1.5).

Lemma 5.2. *Assume that (H_1)–(H_4) hold. Then, for all $k \in \mathbb{N}$,*

$$V_k = H^k\,V_0. \tag{5.23}$$

Proof. With the filtration $(\mathcal{F}_t)$ defined in (2.2), for $k \geq 1$,

$$\begin{aligned}
v_k^{(0)} &= \mathbb{E}[\theta_{k-1}^4\,\ldots\,\theta_1^4\,\mathbb{E}[\theta_k^4\,|\,\mathcal{F}_{k-1}]] \\
&= (\theta^4 + 6\,\theta^2\,\tau_2 + \tau_4)\,v_{k-1}^{(0)} + 4\,\alpha\,(\theta^3 + 3\,\theta\,\tau_2)\,v_{k-1}^{(1)} + 6\,\alpha^2\,(\theta^2+\tau_2)\,v_{k-1}^{(2)} \\
&\qquad + 4\,\alpha^3\,\theta\,v_{k-1}^{(3)} + \alpha^4\,v_{k-1}^{(4)}, \\
v_k^{(1)} &= \mathbb{E}[\theta_{k-1}^4\,\ldots\,\theta_1^4\,\mathbb{E}[\eta_k\,\theta_k^4\,|\,\mathcal{F}_{k-1}]] \\
&= (4\,\theta^3\,\tau_2 + 4\,\theta\,\tau_4)\,v_{k-1}^{(0)} + 4\,\alpha\,(3\,\theta^2\,\tau_2 + \tau_4)\,v_{k-1}^{(1)} + 12\,\alpha^2\,\theta\,\tau_2\,v_{k-1}^{(2)} \\
&\qquad + 4\,\alpha^3\,\tau_2\,v_{k-1}^{(3)}, \\
v_k^{(2)} &= \mathbb{E}[\theta_{k-1}^4\,\ldots\,\theta_1^4\,\mathbb{E}[\eta_k^2\,\theta_k^4\,|\,\mathcal{F}_{k-1}]] \\
&= (\theta^4\,\tau_2 + 6\,\theta^2\,\tau_4 + \tau_6)\,v_{k-1}^{(0)} + 4\,\alpha\,(\theta^3\,\tau_2 + 3\,\theta\,\tau_4)\,v_{k-1}^{(1)} + 6\,\alpha^2\,(\theta^2\,\tau_2+\tau_4)\,v_{k-1}^{(2)} \\
&\qquad + 4\,\alpha^3\,\theta\,\tau_2\,v_{k-1}^{(3)} + \alpha^4\,\tau_2\,v_{k-1}^{(4)}, \\
v_k^{(3)} &= \mathbb{E}[\theta_{k-1}^4\,\ldots\,\theta_1^4\,\mathbb{E}[\eta_k^3\,\theta_k^4\,|\,\mathcal{F}_{k-1}]] \\
&= (4\,\theta^3\,\tau_4 + 4\,\theta\,\tau_6)\,v_{k-1}^{(0)} + 4\,\alpha\,(3\,\theta^2\,\tau_4 + \tau_6)\,v_{k-1}^{(1)} + 12\,\alpha^2\,\theta\,\tau_4\,v_{k-1}^{(2)} \\
&\qquad + 4\,\alpha^3\,\tau_4\,v_{k-1}^{(3)}, \\
v_k^{(4)} &= \mathbb{E}[\theta_{k-1}^4\,\ldots\,\theta_1^4\,\mathbb{E}[\eta_k^4\,\theta_k^4\,|\,\mathcal{F}_{k-1}]] \\
&= (\theta^4\,\tau_4 + 6\,\theta^2\,\tau_6 + \tau_8)\,v_{k-1}^{(0)} + 4\,\alpha\,(\theta^3\,\tau_4 + 3\,\theta\,\tau_6)\,v_{k-1}^{(1)} + 6\,\alpha^2\,(\theta^2\,\tau_4+\tau_6)\,v_{k-1}^{(2)} \\
&\qquad + 4\,\alpha^3\,\theta\,\tau_4\,v_{k-1}^{(3)} + \alpha^4\,\tau_4\,v_{k-1}^{(4)},
\end{aligned}$$

where Table 1 may be read to get the coefficients appearing in the calculations. We reach the matrix formulation $V_k = H\,V_{k-1}$ and the initial value V_1 is obtained *via*

$$\begin{array}{rcl}
v_1^{(0)} &=& \mathbb{E}[\theta_1^4] \;=\; (\theta^4 + 6\,\theta^2\,\tau_2 + \tau_4) + 6\,\alpha^2\,\tau_2\,(\theta^2 + \tau_2) + \alpha^4\,\tau_4,\\
v_1^{(1)} &=& \mathbb{E}[\eta_1\,\theta_1^4] \;=\; (4\,\theta^3\,\tau_2 + 4\,\theta\,\tau_4) + 12\,\alpha^2\,\theta\,\tau_2^2,\\
v_1^{(2)} &=& \mathbb{E}[\eta_1^2\,\theta_1^4] \;=\; (\theta^4\,\tau_2 + 6\,\theta^2\,\tau_4 + \tau_6) + 6\,\alpha^2\,\tau_2\,(\theta^2\,\tau_2 + \tau_4) + \alpha^4\,\tau_2\,\tau_4,\\
v_1^{(3)} &=& \mathbb{E}[\eta_1^3\,\theta_1^4] \;=\; (4\,\theta^3\,\tau_4 + 4\,\theta\,\tau_6) + 12\,\alpha^2\,\theta\,\tau_2\,\tau_4,\\
v_1^{(4)} &=& \mathbb{E}[\eta_1^4\,\theta_1^4] \;=\; (\theta^4\,\tau_4 + 6\,\theta^2\,\tau_6 + \tau_8) + 6\,\alpha^2\,\tau_2\,(\theta^2\,\tau_4 + \tau_6) + \alpha^4\,\tau_4^2.
\end{array}$$

Hence, $V_1 = H\,V_0$. □

Now for all $1 \leq k < \ell$, consider the sequence

$$w_{\ell,k}^{(a)} \;=\; \mathbb{E}[\eta_\ell^a\,\theta_\ell^4\,\ldots\,\theta_{\ell-k+1}^4\,\theta_{\ell-k}^2\,\ldots\,\theta_1^2]$$

where $a \in \{0,\ldots,4\}$, then build

$$W_{\ell,k} = \begin{pmatrix} w_{\ell,k}^{(0)} \\ \vdots \\ w_{\ell,k}^{(4)} \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} \theta^2+\tau_2 & 2\,\alpha\,\theta & \alpha^2 & 0 & 0 \\ 2\,\theta\,\tau_2 & 2\,\alpha\,\tau_2 & 0 & 0 & 0 \\ \theta^2\,\tau_2+\tau_4 & 2\,\alpha\,\theta\,\tau_2 & \alpha^2\,\tau_2 & 0 & 0 \\ 2\,\theta\,\tau_4 & 2\,\alpha\,\tau_4 & 0 & 0 & 0 \\ \theta^2\,\tau_4+\tau_6 & 2\,\alpha\,\theta\,\tau_4 & \alpha^2\,\tau_4 & 0 & 0 \end{pmatrix}.$$

Once again, note that G can be expressed directly from $\{V_0,\ldots,V_4\}$,

$$\left\{\begin{array}{l} G_1 = \theta^2\,V_0 + 2\,\theta\,V_1 + V_2 \\ G_2 = 2\,\alpha\,(\theta\,V_0 + V_1) \\ G_3 = \alpha^2\,V_0 \\ G_4 = 0 \\ G_5 = 0. \end{array}\right. \tag{5.24}$$

Observe also that the upper left-hand 3×3 submatrix of G is precisely M given by (1.4). This argument will be used thereafter to establish that $\rho(G) < 1$.

Lemma 5.3. *Assume that (H$_1$)–(H$_4$) hold. Then, for all* $1 \leq k < \ell$,

$$W_{\ell,k} = H^k\,G^{\ell-k}\,V_0. \tag{5.25}$$

Proof. The calculations are precisely the same as in the proof of Lemmas 5.1 and 5.2. Indeed,

$$W_{\ell,k} = H^k\,U_{\ell-k}$$

where we extend the definition of $U_{k,h}$ in (5.2) to $a \in \{0,\ldots,4\}$, namely

$$U_k = \begin{pmatrix} u_k^{(0)} \\ \vdots \\ u_k^{(4)} \end{pmatrix} = \begin{pmatrix} u_{k,0}^{(0)} \\ \vdots \\ u_{k,0}^{(4)} \end{pmatrix} = U_{k,0}.$$

Then it just remains to investigate the behavior of $u_{\ell-k}$ for $a = 3$ and $a = 4$ using Table 1,

$$u_{\ell-k}^{(3)} \;=\; \mathbb{E}[\theta_{\ell-k-1}^2\,\ldots\,\theta_1^2\,\mathbb{E}[\eta_{\ell-k}^3\,\theta_{\ell-k}^2\,|\,\mathcal{F}_{\ell-k-1}]]$$

$$\begin{aligned} &= 2\,\theta\,\tau_4\,u_{\ell-k-1}^{(0)} + 2\,\alpha\,\tau_4\,u_{\ell-k-1}^{(1)}, \\ u_{\ell-k}^{(4)} &= \mathbb{E}[\theta_{\ell-k-1}^2 \ldots \theta_1^2\,\mathbb{E}[\eta_{\ell-k}^4\,\theta_{\ell-k}^2 \,|\, \mathcal{F}_{\ell-k-1}]] \\ &= (\theta^2\,\tau_4 + \tau_6)\,u_{\ell-k-1}^{(0)} + 2\,\alpha\,\theta\,\tau_4\,u_{\ell-k-1}^{(1)} + \alpha^2\,\tau_4\,u_{\ell-k-1}^{(2)}. \end{aligned}$$

Hence, $U_{\ell-k} = G\,U_{\ell-k-1}$. It is not hard to conclude that, for all $1 \leq k < \ell$,

$$U_{\ell-k} = G^{\ell-k}\,V_0.$$

□

Corollary 5.2. *Assume that (H_1)–(H_4) hold. Then, the fourth-order properties of (X_t) are such that, for all $a \in \{0, \ldots, 4\}$,*

$$\mathbb{E}[\eta_t^a\,X_t^4] < \infty.$$

Proof. For all $t \in \mathbb{Z}$ and $k \geq 1$, denote by

$$\Delta_t = \begin{pmatrix} 1 \\ \eta_t \\ \vdots \\ \eta_t^4 \end{pmatrix} \quad \text{and} \quad P_{t,\,k} = \prod_{i=0}^{k-1} \theta_{t-i} \tag{5.26}$$

with $P_{t,\,0} = 1$. Since (ε_t) and (η_t) are uncorrelated white noises, using the causal representation (2.1) and the same notations as above,

$$\begin{aligned} \mathbb{E}[\Delta_t\,X_t^4] &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{u=0}^{\infty}\sum_{v=0}^{\infty} \mathbb{E}[\Delta_t\,P_{t,\,k}\,P_{t,\,\ell}\,P_{t,\,u}\,P_{t,\,v}\,\varepsilon_{t-k}\,\varepsilon_{t-\ell}\,\varepsilon_{t-u}\,\varepsilon_{t-v}] \\ &= \sigma_4 \sum_{k=0}^{\infty} \mathbb{E}[\Delta_t\,P_{t,\,k}^4] + 6\,\sigma_2^2 \sum_{k=0}^{\infty}\sum_{\ell=k+1}^{\infty} \mathbb{E}[\Delta_t\,P_{t,\,k}^2\,P_{t,\,\ell}^2] \\ &= \sigma_4 \sum_{k=0}^{\infty} V_k + 6\,\sigma_2^2 \sum_{\ell=1}^{\infty} U_\ell + 6\,\sigma_2^2 \sum_{k=1}^{\infty}\sum_{\ell=k+1}^{\infty} W_{\ell,k}. \end{aligned}$$

Then, Lemmas 5.2 and 5.3 together with the strict stationarity of (θ_t) enable to conclude the proof under the assumptions made, since $\rho(G) = \rho(M) < 1$. □

We now return to the proof of Theorem 3.2 and we make the following notations,

$$\begin{aligned} M_n^{(1)} &= \sum_{t=1}^{n} X_{t-1}\left((1 - 2\,\alpha\,\tau_2)\,\varepsilon_t + 2\,\alpha\,\theta\,\eta_t\,\varepsilon_t + 2\,\alpha\,\eta_t^2\,\varepsilon_t\right), \\ M_n^{(2)} &= \sum_{t=1}^{n} X_{t-1}^2\left((1 - 2\,\alpha\,\tau_2 + \alpha\,\theta^2)\,\eta_t + \alpha\,\eta_t^3 + 2\,\alpha\,\theta\,(\eta_t^2 - \tau_2)\right), \\ M_n^{(3)} &= 2\,\alpha^2 \sum_{t=1}^{n} \eta_{t-1}\,X_{t-1}\,\eta_t\,\varepsilon_t, \\ M_n^{(4)} &= \sum_{t=1}^{n} \eta_{t-1}\,X_{t-1}^2\left(2\,\alpha^2\,\theta\,\eta_t + 2\,\alpha^2\,(\eta_t^2 - \tau_2)\right), \end{aligned}$$

$$M_n^{(5)} = \alpha^3 \sum_{t=1}^{n} \eta_{t-1}^2 \, X_{t-1}^2 \, \eta_t,$$

$$M_n^{(6)} = \alpha \sum_{t=1}^{n} \eta_t \, \varepsilon_t^2.$$

Consider the filtration $(\mathcal{F}_n^*)$ given in (5.6), and let

$$\mathcal{M}_n = \begin{pmatrix} M_n^{(1)} \\ \vdots \\ M_n^{(6)} \end{pmatrix}. \tag{5.27}$$

Under our hypotheses, $\mathcal{M}_n$ is a locally square-integrable real vector $(\mathcal{F}_n^*)$–martingale. Once again we will make use of the central limit theorem for vector martingales, as in the proof of Theorem (3.1). On the one hand, we have to study the asymptotic behavior of the predictable quadratic variation of $\mathcal{M}_n$. For all $n \geq 1$, let

$$\langle \mathcal{M} \rangle_n = \sum_{t=1}^{n} \mathbb{E}\big[(\Delta \mathcal{M}_t)(\Delta \mathcal{M}_t)^T \, | \, \mathcal{F}_{t-1}^* \big], \tag{5.28}$$

with $\Delta \mathcal{M}_1 = \mathcal{M}_1$. To simplify the calculations, we introduce some more notations. The second-order moments of the process are defined in (5.9) and its fourth-order moments are called

$$\mathbb{E}[\Delta_t \, X_t^4] = \begin{pmatrix} \delta_0 \\ \vdots \\ \delta_4 \end{pmatrix} = \Delta \tag{5.29}$$

where Δ_t is given in (5.26). We use the strict stationarity to establish, following Corollaries 5.1 and 5.2, that

$$\Delta = (I_5 - H)^{-1} \, (\sigma_2 \, R + \sigma_4 \, V_0) \tag{5.30}$$

in which R is defined from (5.24) as

$$R = 6 \, \lambda_0 \, G_1 + 6 \, \lambda_1 \, G_2 + 6 \, \lambda_2 \, G_3.$$

Now, we are going to show that the asymptotic behavior of $\langle \mathcal{M} \rangle_n / n$ is entirely described by Λ and Δ. By ergodicity,

$$\frac{1}{n} \sum_{t=1}^{n} \Delta_t \, X_t^4 \xrightarrow{\text{a.s.}} \Delta. \tag{5.31}$$

We get back to (5.28). First, there exists constants such that

$$\langle M^{(1)}, M^{(2)} \rangle_n = \sum_{t=1}^{n} X_{t-1}^3 \, \mathbb{E}\big[\big(k_{(1)} + k_{(2)} \, \eta_t + k_{(3)} \, \eta_t^2 \big) \big(k_{(4)} \, \eta_t + k_{(5)} \, \eta_t^3 + k_{(6)} \, (\eta_t^2 - \tau_2) \big) \varepsilon_t \big] = 0$$

under our assumptions. *Via* analogous arguments, it follows that

$$\langle M^{(1)}, M^{(4)} \rangle_n = \langle M^{(1)}, M^{(5)} \rangle_n = \langle M^{(1)}, M^{(6)} \rangle_n = \langle M^{(2)}, M^{(3)} \rangle_n$$

$$= \langle M^{(3)}, M^{(4)} \rangle_n = \langle M^{(3)}, M^{(5)} \rangle_n = \langle M^{(3)}, M^{(6)} \rangle_n = 0.$$

Then we look at nonzero contributions, where we use the constants defined in (A.3) and (A.4). Since the calculations are very similar we only detail the first one,

$$\begin{aligned} \langle M^{(1)} \rangle_n &= \sum_{t=1}^{n} X_{t-1}^2 \, \mathbb{E}\big[\big((1 - 2\,\alpha\,\tau_2)\,\varepsilon_t + 2\,\alpha\,\theta\,\eta_t\,\varepsilon_t + 2\,\alpha\,\eta_t^2\,\varepsilon_t\big)^2\big] \\ &= \sigma_2 \left(1 + 4\,\alpha^2\,(\theta^2\,\tau_2 - \tau_2^2 + \tau_4)\right) \sum_{t=1}^{n} X_{t-1}^2. \end{aligned}$$

To sum up, we obtain

$$\langle \mathcal{M} \rangle_n = K \circ \sum_{t=1}^{n} \begin{pmatrix} X_t^2 & 0 & \eta_t\,X_t^2 & 0 & 0 & 0 \\ 0 & X_t^4 & 0 & \eta_t\,X_t^4 & \eta_t^2\,X_t^4 & X_t^2 \\ \eta_t\,X_t^2 & 0 & \eta_t^2\,X_t^2 & 0 & 0 & 0 \\ 0 & \eta_t\,X_t^4 & 0 & \eta_t^2\,X_t^4 & \eta_t^3\,X_t^4 & \eta_t\,X_t^2 \\ 0 & \eta_t^2\,X_t^4 & 0 & \eta_t^3\,X_t^4 & \eta_t^4\,X_t^4 & \eta_t^2\,X_t^2 \\ 0 & X_t^2 & 0 & \eta_t\,X_t^2 & \eta_t^2\,X_t^2 & 1 \end{pmatrix} + R_n \tag{5.32}$$

where the Hadamard product $\circ$ is used to lighten the formula, and where the remainder R_n is made of isolated terms such that

$$\frac{R_n}{n} \xrightarrow{\text{a.s.}} 0. \tag{5.33}$$

To reach these results, we refer the reader to the explanations following (5.13) since the same methodology has just been applied on $\mathcal{M}_n$. The combination of (5.11), (5.31), (5.32) and (5.33) leads to

$$\frac{\langle \mathcal{M} \rangle_n}{n} \xrightarrow{\text{a.s.}} K \circ \Gamma \tag{5.34}$$

where Γ is given by

$$\Gamma = \begin{pmatrix} \lambda_0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & \delta_0 & 0 & \delta_1 & \delta_2 & \lambda_0 \\ \lambda_1 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & \delta_1 & 0 & \delta_2 & \delta_3 & \lambda_1 \\ 0 & \delta_2 & 0 & \delta_3 & \delta_4 & \lambda_2 \\ 0 & \lambda_0 & 0 & \lambda_1 & \lambda_2 & 1 \end{pmatrix}. \tag{5.35}$$

On the other hand, it is necessary to prove that the Lindeberg's condition is satisfied, namely that for all $\varepsilon > 0$,

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\big[\|\Delta \mathcal{M}_t\|^2 \, \mathbb{I}_{\{\|\Delta \mathcal{M}_t\| \geq \varepsilon \sqrt{n}\}} \,|\, \mathcal{F}_{t-1}^*\big] \xrightarrow{\mathbb{P}} 0 \tag{5.36}$$

as n tends to infinity. The result follows from Corollaries 5.1 and 5.2, together with the same reasoning as the one used to establih (5.16). From (5.34) and (5.36), we deduce that

$$\frac{\mathcal{M}_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, K \circ \Gamma). \tag{5.37}$$

Finally, by a very tedious but straightforward calculation, one can obtain that

$$\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) = \frac{n}{\sum_{t=1}^{n} X_{t-1}^2} \frac{\Omega_6^T \mathcal{M}_n + r_n}{(1 - 2\,\alpha\,\tau_2)\sqrt{n}} \tag{5.38}$$

where $\Omega_6^T = (1\ 1\ 1\ 1\ 1\ 1)$ and $r_n = o(\sqrt{n})$ a.s. from (5.11) and (5.31). It remains to apply Slutsky's lemma to conclude that

$$\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \omega^2)$$

with

$$\omega^2 = \frac{\Omega_6^T (K \circ \Gamma)\, \Omega_6}{\lambda_0^2\,(1 - 2\,\alpha\,\tau_2)^2} \tag{5.39}$$

using the whole notations above. □

5.6. **Proof of Theorem 3.3.** Letting $V_n = \sqrt{n}\, I_6$, such a sequence obviously satisfies the regular growth conditions of [8]. Keeping the notations of (5.27), we have studied the hook of $\mathcal{M}_n$ in (5.34) and Lindeberg's condition is already fulfilled in (5.36), it only remains to check that

$$\frac{[\mathcal{M}]_n - \langle \mathcal{M} \rangle_n}{n} \xrightarrow{\text{a.s.}} 0 \tag{5.40}$$

where

$$[\mathcal{M}]_n = \sum_{t=1}^{n} (\Delta \mathcal{M}_t)(\Delta \mathcal{M}_t)^T$$

is the total variation of $\mathcal{M}_n$, to apply Thm. 2.1 of [8]. To be precise with the required hypotheses, note that (5.36) also holds almost surely, by ergodicity. But (5.40) is an immediate consequence of the ergodicity of the increments. Thus,

$$\frac{1}{6\,\ln n} \sum_{t=1}^{n} \left[1 - \left(\frac{t}{t+1}\right)^6\right] \frac{\mathcal{M}_t\, \mathcal{M}_t^T}{t} \xrightarrow{\text{a.s.}} K \circ \Gamma$$

and, after simplifications,

$$\frac{1}{\ln n} \sum_{t=1}^{n} \frac{\mathcal{M}_t\, \mathcal{M}_t^T}{t^{\,2}} \xrightarrow{\text{a.s.}} K \circ \Gamma. \tag{5.41}$$

The remainder r_n in (5.38) is a long linear combination of isolated terms, we detail here the treatment of the largest one which takes the form of $\eta_{n-1}^2\, X_{n-1}^2\, \eta_n$. Corollary 5.2 implies, for $a = 4$ and *via* the ergodic theorem, that

$$\frac{1}{n} \sum_{t=1}^{n} \eta_{t-1}^4\, X_{t-1}^4\, \eta_t^2 \xrightarrow{\text{a.s.}} \delta_4\, \tau_2,$$

which in turn leads to

$$\frac{\eta_{n-1}^4\, X_{n-1}^4\, \eta_n^2}{n} \xrightarrow{\text{a.s.}} 0 \quad \text{so that} \quad \frac{\eta_{n-1}^4\, X_{n-1}^4\, \eta_n^2}{n^2} = o(n^{-1}) \ \text{ a.s.}$$

It follows that

$$\sum_{t=1}^{n} \frac{\eta_{t-1}^4 \, X_{t-1}^4 \, \eta_t^2}{t^2} = o\bigg(\sum_{t=1}^{n} \frac{1}{t} \bigg) = o(\ln n) \quad \text{a.s.}$$

By extrapolation, treating similarly all residual terms,

$$\frac{1}{\ln n} \sum_{t=1}^{n} \frac{r_t^2}{t^2} \xrightarrow{\text{a.s.}} 0. \tag{5.42}$$

It remains to combine these results to get

$$\begin{aligned} \frac{(1-2\,\alpha\,\tau_2)^2}{\ln n} \sum_{t=1}^{n} \big(\widehat{\theta}_t - \theta^*\big)^2 &= \frac{1}{\ln n} \sum_{t=1}^{n} \frac{\Omega_6^T \mathcal{M}_t \, \mathcal{M}_t^T \Omega_6}{S_{t-1}^2} + \frac{1}{\ln n} \sum_{t=1}^{n} \frac{r_t^2}{S_{t-1}^2} \\ &\quad + \frac{2}{\ln n} \sum_{t=1}^{n} \frac{\Omega_6^T \mathcal{M}_t \, r_t}{S_{t-1}^2} \end{aligned}$$

where

$$S_n = \sum_{t=0}^{n} X_t^2 \quad \text{satisfies} \quad \frac{S_n}{n} \xrightarrow{\text{a.s.}} \lambda_0. \tag{5.43}$$

Using Cauchy-Schwarz inequality, the cross-term is shown to be negligible. From (5.39), (5.41), (5.42) and the previous remark,

$$\frac{1}{\ln n} \sum_{t=1}^{n} \big(\widehat{\theta}_t - \theta^*\big)^2 \xrightarrow{\text{a.s.}} \frac{\Omega_6^T (K \circ \Gamma) \, \Omega_6}{\lambda_0^2 \, (1-2\,\alpha\,\tau_2)^2} = \omega^2$$

which concludes the first part of the proof and follows from Toeplitz lemma applied in the right-hand side of the decomposition. The rate of convergence of $\widehat{\theta}_n$ is easier to handle. As a matter of fact, we have already seen that $\mathcal{M}_n$ is a vector $(\mathcal{F}_n^*)$–martingale having ergodic and stationary increments. So,

$$\mathcal{N}_n = \Omega_6^T \mathcal{M}_n \tag{5.44}$$

is a scalar $(\mathcal{F}_n^*)$–martingale having the same incremental properties, and our hypotheses guarantee that $\mathbb{E}[(\Delta \mathcal{N}_1)^2] = \Omega_6^T (K \circ \Gamma) \, \Omega_6 < \infty$. The main theorem of [20] enables to infer that

$$\limsup_{n \to +\infty} \frac{\mathcal{N}_n}{\sqrt{2\,n\,\ln\ln n}} = \sqrt{\Omega_6^T (K \circ \Gamma) \, \Omega_6} \quad \text{a.s.} \tag{5.45}$$

and

$$\liminf_{n \to +\infty} \frac{\mathcal{N}_n}{\sqrt{2\,n\,\ln\ln n}} = -\sqrt{\Omega_6^T (K \circ \Gamma) \, \Omega_6} \quad \text{a.s.} \tag{5.46}$$

replacing $\mathcal{N}_n$ by $-\mathcal{N}_n$. Thus, once again exploiting (5.38),

$$\begin{aligned} \limsup_{n \to +\infty} \sqrt{\frac{n}{2\,\ln\ln n}} \, \big(\widehat{\theta}_n - \theta^*\big) &= \frac{1}{\lambda_0 \, (1-2\,\alpha\,\tau_2)} \limsup_{n \to +\infty} \frac{\mathcal{N}_n + r_n}{\sqrt{2\,n\,\ln\ln n}} \\ &= \omega \quad \text{a.s.} \end{aligned}$$

using (5.45) and the fact that $r_n = o(\sqrt{n})$ a.s. The symmetric result is reached from (5.46) and the proof is complete. □

5.7. **Proof of Theorem 4.1.** One shall prove this result in two steps. First, we will identify the covariance Σ such that

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}_n - \theta^* \\ \widehat{\vartheta}_n - \vartheta^*\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma) \tag{5.47}$$

where $\widehat{\theta}_n$ and $\widehat{\vartheta}_n$ are given in (4.1), $\theta^* = \rho_X(1)$ is the limiting value of $\widehat{\theta}_n$ deeply investigated up to this point and

$$\vartheta^* = \rho_X(2) = \frac{\theta^2 + \alpha\,\tau_2\,(1 - 2\,\alpha\,\tau_2)}{1 - 2\,\alpha\,\tau_2}.$$

Then we will translate the result to the new estimates (4.3) *via* the Delta method. Of course the first step being very close to the proof of Theorem 3.2, we only give an outline of the calculations. The second-order lag in $\widehat{\vartheta}_n$ gives a new scalar $(\mathcal{F}_n^*)$–martingale contribution that we will define as

$$\begin{aligned}\mathcal{L}_n &= \alpha\sum_{t=1}^{n} X_{t-1}\,\eta_t\,\varepsilon_t + \sum_{t=1}^{n} X_{t-1}^2\left(\alpha\,\theta\,\eta_t + \alpha\,(\eta_t^2 - \tau_2)\right) \\ &\quad + \alpha^2\sum_{t=1}^{n}\eta_{t-1}\,X_{t-1}^2\,\eta_t + \sum_{t=2}^{n} X_{t-2}\,\varepsilon_t + \sum_{t=2}^{n} X_{t-2}\,\varepsilon_{t-1}\,\eta_t \\ &\quad + \theta\sum_{t=2}^{n} X_{t-2}^2\,\eta_t + \sum_{t=2}^{n} X_{t-2}^2\,\eta_{t-1}\,\eta_t + \alpha\sum_{t=2}^{n}\eta_{t-2}\,X_{t-2}^2\,\eta_t\end{aligned} \tag{5.48}$$

which follows from a very tedious development of $\sum_{t=2}^{n} X_{t-2}X_t$. An exhaustive expansion of $\widehat{\vartheta}_n - \vartheta^*$ leads to

$$\left(\widehat{\vartheta}_n - \vartheta^*\right)S_{n-2} = \theta^*\,\Omega_6^T\,\mathcal{M}_n + \mathcal{L}_n + s_n$$

where $\mathcal{M}_n$ is given in (5.27), S_n in (5.43), $\Omega_6^T = (1\ 1\ 1\ 1\ 1\ 1)$ and s_n is made of isolated terms, each one being $o(\sqrt{n})$ a.s. as soon as the process has fourth-order moments, *i.e.* under (H_4). Combined with (5.38),

$$\sqrt{n}\begin{pmatrix}\widehat{\theta}_n - \theta^* \\ \widehat{\vartheta}_n - \vartheta^*\end{pmatrix} = \frac{A_n}{\sqrt{n}}\begin{pmatrix}\mathcal{M}_n \\ \mathcal{L}_n\end{pmatrix} + T_n \tag{5.49}$$

where

$$A_n = \begin{pmatrix}\frac{n}{S_{n-1}} & \frac{\Omega_6^T}{1-2\,\alpha\,\tau_2} & 0 \\ \frac{n}{S_{n-2}} & \frac{\theta\,\Omega_6^T}{1-2\,\alpha\,\tau_2} & \frac{n}{S_{n-2}}\end{pmatrix} \xrightarrow{\text{a.s.}} A = \begin{pmatrix}\frac{\Omega_6^T}{\lambda_0\,(1-2\,\alpha\,\tau_2)} & 0 \\ \frac{\theta\,\Omega_6^T}{\lambda_0\,(1-2\,\alpha\,\tau_2)} & \frac{1}{\lambda_0}\end{pmatrix} \tag{5.50}$$

are matrices of size 2×7 and $T_n = o(1)$ a.s. We have to study the hook of this new vector $(\mathcal{F}_n^*)$–martingale. First, $\langle\mathcal{M}\rangle_n$ is already treated in (5.34). For the cross-term and the last one, we need more notations. Let

$$\mu_{a,b,c,p,q} = \mathbb{E}[\eta_{t-1}^a\,\eta_t^b\,\varepsilon_t^c\,X_{t-1}^p\,X_t^q] \tag{5.51}$$

and observe that $\mu_{0,b,0,0,2} = [\Lambda]_{b+1}$ in (5.9) for $b \in \{0,1,2\}$ and that $\mu_{0,b,0,0,4} = [\Delta]_{b+1}$ in (5.29) for $b \in \{0,\ldots,4\}$. Then, it can be seen *via* analogous arguments as usual relying on ergodicity and negligible isolated terms, that

$$\frac{\langle \mathcal{M}, \mathcal{L} \rangle_n}{n} \xrightarrow{\text{a.s.}} (L \circ \Upsilon)\, \Omega_6 \tag{5.52}$$

where L is defined in (A.6) and Υ is given by

$$\Upsilon = \begin{pmatrix} \theta^* \lambda_0 & \lambda_0 & 0 & 0 & 0 & 0 \\ \delta_0 & \delta_1 & \mu_{0,0,0,2,2} & \mu_{0,1,0,2,2} & \mu_{1,0,0,2,2} & \mu_{0,0,1,1,2} \\ \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ \delta_1 & \delta_2 & \mu_{0,1,0,2,2} & \mu_{0,2,0,2,2} & \mu_{1,1,0,2,2} & \mu_{0,1,1,1,2} \\ \delta_2 & \delta_3 & \mu_{0,2,0,2,2} & \mu_{0,3,0,2,2} & \mu_{1,2,0,2,2} & \mu_{0,2,1,1,2} \\ \lambda_0 & \lambda_1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5.53}$$

Finally, we have

$$\begin{aligned} \frac{\langle \mathcal{L} \rangle_n}{n} \xrightarrow{\text{a.s.}} \ell \ = \ & m_{(1)}\, \lambda_0 + m_{(2)}\, \delta_0 + m_{(3)}\, \delta_1 + m_{(4)}\, \delta_2 + \theta\, m_{(5)}\, \mu_{0,0,0,2,2} \\ & + \alpha\, m_{(5)}\, \mu_{1,0,0,2,2} + (1+\alpha)\, m_{(5)}\, \mu_{0,1,0,2,2} + m_{(5)}\, \mu_{0,0,1,1,2} \\ & + m_{(6)}\, \mu_{0,2,0,2,2} + \alpha\, m_{(6)}\, \mu_{1,1,0,2,2} + m_{(6)}\, \mu_{0,1,1,1,2} \end{aligned} \tag{5.54}$$

where the constants are detailed in (A.7). This last convergence, together with (5.52), (5.34) and their related notations, implies

$$\frac{1}{n} \left\langle \begin{pmatrix} \mathcal{M} \\ \mathcal{L} \end{pmatrix} \right\rangle_n \xrightarrow{\text{a.s.}} \Sigma_{\text{ML}} = \begin{pmatrix} K \circ \Gamma & (L \circ \Upsilon)\, \Omega_6 \\ \Omega_6^T (L \circ \Upsilon)^T & \ell \end{pmatrix}. \tag{5.55}$$

Lindeberg's condition is clearly fulfilled and Slutsky's lemma applied on the relation (5.49), taking into account the asymptotic normality of the martingale and the remarks that follow (5.49), enables to identify Σ in (5.47) as

$$\Sigma = A\, \Sigma_{\text{ML}} A^T \tag{5.56}$$

where A is given in (5.50). This ends the first part of the proof.

Remark 5.1. *It is important to note that, despite the complex structure of Σ, it only depends on the parameters and can be computed explicitely. Indeed, it is easy to see that all coefficients $\mu_{a,b,c,p,q}$ in Σ_{ML} exist under our hypotheses, exploiting the fourth-order moments of the process. We can compute each of them using the same lines as in our previous technical lemmas.*

Consider now the mapping f in (4.2) whose Jacobian matrix is

$$\nabla f(x,y) = \begin{pmatrix} \frac{(1-2y)\,(1+2x^2)}{(1-2x^2)^2} & \frac{-2x}{1-2x^2} \\ \frac{-2x\,(1-2y)}{(1-2x^2)^2} & \frac{1}{1-2x^2} \end{pmatrix}.$$

The couple of estimates (4.3) therefore satisfies

$$\sqrt{n} \begin{pmatrix} \widetilde{\theta}_n - \theta \\ \widetilde{\gamma}_n - \gamma \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \nabla^T f(\theta^*, \vartheta^*)\, \Sigma\, \nabla f(\theta^*, \vartheta^*))$$

by application of the Delta method, the pathological cases $\theta^* = \pm\frac{1}{\sqrt{2}}$ being excluded from the study. □

Appendix

This appendix is devoted to the numerous constants of the study, for greater clarity. The first of them are given by

$$\text{(A.1)} \qquad \left\{ \begin{array}{lcl} \bar{k}_{(1)} & = & (1+\alpha\,\theta)^2\,\tau_2 + \alpha^2\,(\tau_4 - \tau_2^2) \\ \bar{k}_{(1-2)} & = & \alpha^2\,(1+\alpha\,\theta)\,\tau_2 \\ \bar{k}_{(2)} & = & \alpha^4\,\tau_2 \\ \bar{k}_{(3)} & = & (1+\alpha^2\,\tau_2)\,\sigma_2 \end{array} \right.$$

and serve to build the matrix

$$\text{(A.2)} \qquad \bar{K} = \begin{pmatrix} \bar{k}_{(1)} & \bar{k}_{(1-2)} & 0 \\ \bar{k}_{(1-2)} & \bar{k}_{(2)} & 0 \\ 0 & 0 & \bar{k}_{(3)} \end{pmatrix}.$$

We also define

$$\text{(A.3)} \qquad \left\{ \begin{array}{lcl} k_{(1)} & = & \sigma_2\,(1+4\,\alpha^2\,(\theta^2\,\tau_2 - \tau_2^2 + \tau_4)) \\ k_{(1-3)} & = & 4\,\alpha^3\,\theta\,\tau_2\,\sigma_2 \\ k_{(2)} & = & (1-2\,\alpha\,\tau_2 + \alpha\,\theta^2)\,(2\,\alpha\,\tau_4 + \tau_2\,(1-2\,\alpha\,\tau_2 + \alpha\,\theta^2)) \\ & & \quad + \alpha^2\,(\tau_6 + 4\,\theta^2\,(\tau_4 - \tau_2^2)) \\ k_{(2-4)} & = & 2\,\alpha^2\,\theta\,\tau_2\,(1+\alpha\,\theta^2 - 4\,\alpha\,\tau_2) + 6\,\alpha^3\,\theta\,\tau_4 \\ k_{(2-5)} & = & \alpha^3\,(\alpha\,\tau_4 + \tau_2\,(1-2\,\alpha\,\tau_2 + \alpha\,\theta^2)) \\ k_{(2-6)} & = & \alpha\,\sigma_2\,(\alpha\,\tau_4 + \tau_2\,(1-2\,\alpha\,\tau_2 + \alpha\,\theta^2)) \\ k_{(3)} & = & 4\,\alpha^4\,\tau_2\,\sigma_2 \\ k_{(4)} & = & 4\,\alpha^4\,(\theta^2\,\tau_2 - \tau_2^2 + \tau_4) \\ k_{(4-5)} & = & 2\,\alpha^5\,\theta\,\tau_2 \\ k_{(4-6)} & = & 2\,\alpha^3\,\theta\,\tau_2\,\sigma_2 \\ k_{(5)} & = & \alpha^6\,\tau_2 \\ k_{(5-6)} & = & \alpha^4\,\tau_2\,\sigma_2 \\ k_{(6)} & = & \alpha^2\,\tau_2\,\sigma_4 \end{array} \right.$$

that we put in the matrix form

$$\text{(A.4)} \qquad K = \begin{pmatrix} k_{(1)} & 0 & k_{(1-3)} & 0 & 0 & 0 \\ 0 & k_{(2)} & 0 & k_{(2-4)} & k_{(2-5)} & k_{(2-6)} \\ k_{(1-3)} & 0 & k_{(3)} & 0 & 0 & 0 \\ 0 & k_{(2-4)} & 0 & k_{(4)} & k_{(4-5)} & k_{(4-6)} \\ 0 & k_{(2-5)} & 0 & k_{(4-5)} & k_{(5)} & k_{(5-6)} \\ 0 & k_{(2-6)} & 0 & k_{(4-6)} & k_{(5-6)} & k_{(6)} \end{pmatrix}.$$

Moreover, we have to consider

$$\text{(A.5)}\qquad \left\{\begin{array}{rcl} \ell'_{(1)} &=& \sigma_2 \\ \ell_{(1)} &=& 2\,\alpha^2\,\theta\,\tau_2\,\sigma_2 \\ \ell'_{(2)} &=& \alpha\,\theta\,(\tau_2\,(1-2\,\alpha\,\tau_2+\alpha\,\theta^2)-\alpha\,(2\,\tau_2^2-3\,\tau_4)) \\ \ell_{(2)} &=& \alpha\,\tau_4+\tau_2\,(1-2\,\alpha\,\tau_2+\alpha\,\theta^2) \\ \ell_{(3)} &=& 2\,\alpha^3\,\tau_2\,\sigma_2 \\ \ell'_{(4)} &=& 2\,\alpha^3\,(\theta^2\,\tau_2-\tau_2^2+\tau_4) \\ \ell_{(4)} &=& 2\,\alpha^2\,\theta\,\tau_2 \\ \ell_{(5)} &=& \alpha^4\,\tau_2 \\ \ell_{(6)} &=& \alpha\,\tau_2\,\sigma_2\,(1+\alpha) \end{array}\right.$$

in the matrix form

$$\text{(A.6)}\qquad L=\begin{pmatrix} \ell'_{(1)} & \ell_{(1)} & 0 & 0 & 0 & 0 \\ \ell'_{(2)} & \alpha^2\,\ell_{(2)} & \theta\,\ell_{(2)} & \ell_{(2)} & \alpha\,\ell_{(2)} & \ell_{(2)} \\ \ell_{(3)} & 0 & 0 & 0 & 0 & 0 \\ \ell'_{(4)} & \alpha^2\,\ell_{(4)} & \theta\,\ell_{(4)} & \ell_{(4)} & \alpha\,\ell_{(4)} & \ell_{(4)} \\ \alpha\,\theta\,\ell_{(5)} & \alpha^2\,\ell_{(5)} & \theta\,\ell_{(5)} & \ell_{(5)} & \alpha\,\ell_{(5)} & \ell_{(5)} \\ \theta\,\ell_{(6)} & \alpha\,\ell_{(6)} & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We conclude by a last set of constants,

$$\text{(A.7)}\qquad \left\{\begin{array}{rcl} m_{(1)} &=& \sigma_2\,(1+\tau_2\,(1+\alpha^2)) \\ m_{(2)} &=& \theta^2\,(1+\alpha^2)\,\tau_2+(1-\alpha^2)\,\tau_2^2+\alpha^2\,\tau_4 \\ m_{(3)} &=& 2\,\alpha\,\theta\,(1+\alpha^2)\,\tau_2 \\ m_{(4)} &=& \alpha^2\,(1+\alpha^2)\,\tau_2 \\ m_{(5)} &=& 2\,\alpha\,\theta\,\tau_2 \\ m_{(6)} &=& 2\,\alpha^2\,\tau_2. \end{array}\right.$$

References

[1] J. Anděl. Autoregressive series with random parameters. *Math. Operationsforsch. Statist.*, 7-5:735–741, 1976.

[2] A. Aue and L. Horváth. Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Stat. Sinica.*, 21:973–999, 2011.

[3] A. Aue, L. Horváth, and J. Steinebach. Estimation in random coefficient autoregressive models. *J. Time. Ser. Anal.*, 27-1:61–76, 2006.

[4] I. Berkes, L. Horváth, and S. Ling. Estimation in nonstationary random coefficient autoregressive models. *J. Time. Ser. Anal.*, 30-4:395–416, 2009.

[5] P. Billingsley. The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.*, 12:788–792, 1961.

[6] A. Brandt. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.*, 18:211–220, 1986.

[7] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods. Second Edition.* Springer Series in Statistics. Springer-Verlag, New-York, 1991.

[8] F. Chaabane and F. Maaouia. Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab. Stat.*, 4:137–189, 2000.

[9] M. Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics, New York.* Springer-Verlag, Berlin, 1997.

[10] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, New-York, 1985.
[11] S. Y. Hwang and I. V. Basawa. Explosive random-coefficient ar(1) processes and related asymptotics for least-squares estimation. *J. Time. Ser. Anal.*, 26-6:807–824, 2005.
[12] S. Y. Hwang, I. V. Basawa, and T. Y. Kim. Least squares estimation for critical random coefficient first-order autoregressive processes. *Stat. Probab. Lett.*, 76:310–317, 2006.
[13] U. Jürgens. The estimation of a random coefficient AR(1) process under moment conditions. *Statist. Hefte.*, 26:237–249, 1985.
[14] A. Koubkovà. First-order autoregressive processes with time-dependent random parameters. *Kybernetika.*, 18-5:408–414, 1982.
[15] D. F. Nicholls and B. G. Quinn. The estimation of multivariate random coefficient autoregressive models. *J. Multivar. Anal.*, 11:544–555, 1981.
[16] D. F. Nicholls and B. G. Quinn. Multiple autoregressive models with random coefficients. *J. Multivar. Anal.*, 11:185–198, 1981.
[17] D. F. Nicholls and B. G. Quinn. *Random Coefficient Autoregressive Models: An Introduction*, volume 11 of *Lecture Notes in Statistics*. Springer-Verlag, New-York, 1982.
[18] P. M. Robinson. Statistical inference for a random coefficient autoregressive model. *Scand. J. Stat.*, 5-3:163–168, 1978.
[19] A. Schick. $\sqrt{n}$–consistent estimation in a random coefficient autoregressive model. *Austral. J. Statist.*, 38-2:155–160, 1996.
[20] W. F. Stout. The Hartman-Wintner law of the iterated logarithm for martingales. *Ann. Math. Stat.*, 41-6:2158–2160, 1970.
[21] W. F. Stout. *Almost sure convergence*, volume 24 of *Probability and Mathematical Statistics*. Academic Press, New-York-London, 1974.
[22] M. Taniguchi and Y. Kakizawa. *Asymptotic Theory of Statistical Inference for Time Series*. Springer Series in Statistics. Springer, New-York, 2000.

E-mail address: frederic.proia@univ-angers.fr
E-mail address: marius.soltane.etu@univ-lemans.fr

Laboratoire Angevin de REcherche en MAthématiques (LAREMA), CNRS, Université d'Angers, Université Bretagne Loire. 2 Boulevard Lavoisier, 49045 Angers cedex 01, France.

Laboratoire Manceau de Mathématiques, Le Mans Université, Avenue O. Messiaen, 72085 Le Mans cedex 9, France.

Bibliographie

Brockwell, P. J. et Davis, R. A. (2006). *Time series : theory and methods*. Springer Series in Statistics. Springer, New York.

Buchmann, B. et Chan, N. H. (2013). Unified asymptotic theory for nearly unstable AR(p) processes. *Stoch. Proc. Appl.*, 123:952–985.

Chan, N. H. et Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.*, 16:367–401.

Dembo, A. et Zeitouni, O. (1998). *Large Deviations Techniques and Applications (Second Edition)*, volume 38 de *Applications of Mathematics*. Springer.

Lai, T. L. et Wei, C. Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.*, 13:1–23.

Miao, Y., Wang, Y. et Yang, G. (2015). Moderate deviation principles for empirical covariance in the neighbourhood of the unit root. *Scand. J. Stat.*, 42:234–255.

Nicholls, D. F. et Quinn, B. G. (1981). The estimation of multivariate random coefficient autoregressive models. *J. Multivar. Anal.*, 11:544–555.

Phillips, P. C. B. et Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *J. Econometrics.*, 136:115–130.

Proïa, F. (2020). Moderate deviations in a class of stable but nearly unstable processes. *J. Stat. Plan. Infer.*, 208:66–81.

Proïa, F. et Soltane, M. (2018). A test of correlation in the random coefficients of an autoregressive process. *Math. Meth. Stat.*, 26(2):119–144.

Proïa, F. et Soltane, M. (2021). Comments on the presence of serial correlation in the random coefficients of an autoregressive process. *Stat. Probab. Lett.*, 170.

Worms, J. (1999). Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.*, 4:1–28.

Chapitre 2

Modèles graphiques partiels

Dans ce deuxième chapitre, nous allons nous intéresser aux modèles graphiques partiels gaussiens (PGGM), dans un contexte de grande dimension. Nous explorerons dans un premier temps une procédure d'estimation par maximum de vraisemblance pénalisée avant de nous focaliser, dans un second temps, sur une contrepartie bayésienne. Nous donnerons à cet égard le contenu explicite de deux articles, un par section, fruit de travaux effectués en collaboration avec E. Okome Obiang et P. Jézéquel. Mais avant cela, il semble important de fournir quelques détails techniques sur le principe des PGGMs. Supposons que $Z \sim \mathcal{N}_d(0, \Sigma)$ avec $\Sigma \in \mathbb{S}^d_{++}$. Un tel vecteur gaussien admet Σ pour matrice de covariance et $\Omega = \Sigma^{-1}$ pour matrice de précision. L'estimation de Ω est à la base de l'inférence dans les modèles graphiques gaussiens, voir par exemple Maathuis *et al.* (2018) pour un tour d'horizon complet sur le sujet ou encore (Giraud, 2014, Chap. 7) pour un premier aperçu. L'intérêt de travailler sur Ω plutôt que sur Σ repose sur une propriété très forte des vecteurs gaussiens, qui stipule que

$$\mathbb{C}\mathrm{orr}(Z_i, Z_j \mid Z_{\neq i,\, j}) = -\frac{\Omega_{ij}}{\sqrt{\Omega_{ii}\,\Omega_{jj}}} \tag{2.1}$$

pour tous $1 \leqslant i, j \leqslant d$, en d'autres termes que la corrélation partielle entre deux coordonnées du vecteur gaussien Z se lit directement dans sa matrice de précision. En particulier, on voit que dans un tel contexte gaussien, l'indépendance conditionnelle entre Z_i et Z_j est équivalente à $\Omega_{ij} = 0$. Les liens directs entre les composantes de Z peuvent donc se représenter selon une structure de graphe dans lequel les nœuds sont les composantes et deux composantes sont reliées par une arête si l'emplacement correspondant dans Ω n'est pas nul (voir la partie gauche de la Figure 2.1, très schématique), d'où l'importance cruciale pour une procédure statistique de récupérer le support de Ω, et cela passe par sa capacité à imposer de la sparsité (par exemple avec de la pénalisation dans une approche fréquentiste, ou grâce à une stratégie *spike-and-slab* en bayésien, puisque ce sont les cas de figure qui vont nous intéresser par la suite). Plaçons-nous désormais dans un contexte de régression et posons $Z = (Y, X)$ avec $Y \in \mathbb{R}^q$, $X \in \mathbb{R}^p$ et donc $d = q + p$. Comme évoqué dans l'introduction, cela revient à considérer qu'il existe une relation linéaire de la forme

$$Y = B^T X + E \quad \text{avec} \quad E \sim \mathcal{N}_q(0, R) \tag{2.2}$$

et comme nouveaux paramètres $B = -\Delta^T \Omega_y^{-1}$ et $R = \Omega_y^{-1}$ où Δ et Ω_y sont extraits de Ω selon la décomposition en blocs

$$\Omega = \begin{pmatrix} \Omega_y & \Delta \\ \Delta^T & \Omega_x \end{pmatrix}. \tag{2.3}$$

L'estimation du couple (Ω_y, Δ) est alors une alternative qui possède un avantage primordial pour l'interprétation statistique : on a accès à travers Δ aux corrélations partielles, et donc aux liens directs entre les prédicteurs et les réponses. Cela n'est pas possible par l'estimation seule de B, qui peut contenir des liens indirects (en raison par exemple d'une corrélation forte entre les réponses (tout du moins pour $q > 1$), lorsque Ω_y n'est pas diagonale). Dans un PGGM, on cherchera donc à estimer conjointement Δ et Ω_y en lieu et place de B. Par ailleurs et comme nous l'avons détaillé en section introductive, dans un contexte de régression en grande dimension par rapport à p, extraire ces estimations de taille $O(p)$ de celle de Ω induit un biais conséquent car cette dernière étant elle-même de très grande dimension, en $O(p^2)$, son estimation sera nécessairement imprécise (en raison d'effets de pénalisations, de *shrinkage*, etc.). D'où l'intérêt de développer des méthodes d'estimation qui parviennent à se focaliser sur Δ et Ω_y, en laissant Ω_x de côté. À titre d'exemple, considérons un vecteur gaussien $(Y_1, Y_2, X_1, X_2, X_3)$ dans lequel il existe les corrélations partielles $X_1, X_2 \leftrightarrow Y_1$ et $X_3 \leftrightarrow Y_1, Y_2$ avec $Y_1 \leftrightarrow Y_2$ (à rapprocher d'un modèle linéaire dans lequel X_1 et X_2 expliquent Y_1, X_3 explique Y_1 et Y_2 et les réponses Y_1 et Y_2 sont corrélées). Alors, ses matrices de covariance et de précision prendraient la forme

$$\Sigma = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix} \quad \text{et} \quad \Omega = \begin{pmatrix} * & * & * & * & * \\ * & * & 0 & 0 & * \\ * & 0 & * & * & * \\ * & 0 & * & * & * \\ * & * & * & * & * \end{pmatrix} \tag{2.4}$$

et on pourra trouver sur la partie droite de la Figure 2.1 une schématisation du modèle graphique partiel qui en découlerait (code couleurs compris). Notons bien ici que les liens potentiels entre les covariables (en gris ci-dessus) ne nous intéressent pas, on souhaite laisser Ω_x hors de l'étude.

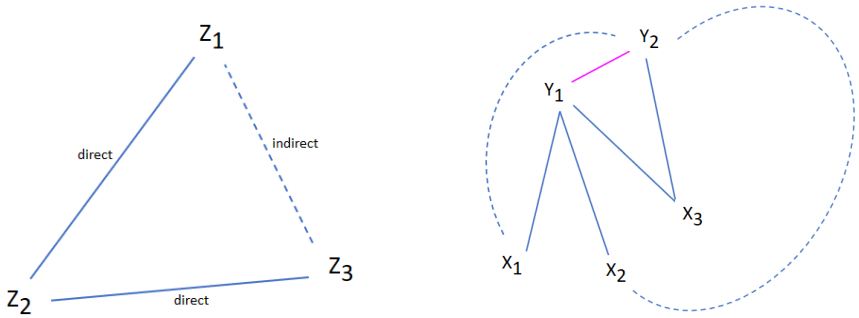


FIGURE 2.1 – Schématisation d'un modèle graphique et des corrélations partielles dans un vecteur gaussien (Z_1, Z_2, Z_3) avec $Z_1 \leftrightarrow Z_2$ et $Z_2 \leftrightarrow Z_3$ mais $Z_1 \not\leftrightarrow Z_3$, à gauche. Les liens directs sont en trait plein, les liens indirects en pointillés. Schématisation du modèle graphique partiel découlant de l'exemple cité ci-dessus avec les liens entre les covariables et les réponses (bleu) et les liens entre les réponses (violet).

2.1 Approche par vraisemblance pénalisée

Maximiser la vraisemblance pénalisée du modèle graphique gaussien sur un échantillon de taille n revient à minimiser sur $\mathbb{S}_{++}^d$ l'objectif

$$L_n(\Omega) = -\ln\det(\Omega) + \operatorname{tr}(S_n\,\Omega) + \lambda\,\mathrm{pen}(\Omega) \tag{2.5}$$

où S_n est la matrice de covariance empirique des observations et $\mathrm{pen}(\Omega)$ est une fonction de pénalisation munie de son paramètre de régularisation $\lambda \geqslant 0$, généralement de type ℓ_1 à des fins de sparsité. Le choix $\mathrm{pen}(\Omega) = |\Omega|_1$ correspond au *Graphical Lasso* bien connu de Friedman *et al.* (2008) mais on rencontre aussi fréquemment $\mathrm{pen}(\Omega) = |\Omega|_1^-$ qui revient à ne pas pénaliser les éléments diagonaux. Pour pallier le problème évoqué à la fin du paragraphe précédent, Yuan et Zhang (2014) montrent que l'on peut faire disparaître Ω_x par une étape préalable d'optimisation et que dans un PGGM pénalisé, on se ramène à la minimisation de l'objectif

$$\begin{aligned} L_n(\Omega_y, \Delta) \;=\; & -\ln\det(\Omega_y) + \operatorname{tr}(S_{n,y}\,\Omega_y) + 2\operatorname{tr}(S_{n,yx}^T\,\Delta) \\ & + \operatorname{tr}(S_{n,x}\,\Delta^T\Omega_y^{-1}\Delta) + \lambda\,\mathrm{pen}(\Omega_y) + \mu\,\mathrm{pen}(\Delta) \end{aligned} \tag{2.6}$$

où $S_{n,y}$, $S_{n,x}$ et $S_{n,yx}$ désignent respectivement la variance empirique des réponses, celle des prédicteurs et leur covariance empirique. Chiquet *et al.* (2017) proposent de remplacer la pénalisation sur Ω_y par une seconde pénalisation sur Δ impliquant un terme de la forme $\operatorname{tr}(\Delta L\,\Delta^T\Omega_y^{-1})$ pour une matrice $L \in \mathbb{S}_+^p$ à choisir, et ayant un effet structurant de par sa construction. Par exemple avec

$$L = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix} \tag{2.7}$$

qui représente la matrice des différences finies du premier ordre, la présence de $\Delta L\,\Delta^T$ induira une pénalisation dans les différences entre deux valeurs voisines (en ligne) de Δ, agissant de façon similaire à la version *fused* du Lasso, et favorisera la sélection de segments de prédicteurs plutôt que de prédicteurs isolés. Pour illustrer cela, dans le cas où R est diagonale avec $R = \mathrm{diag}(\sigma_1^2, \dots, \sigma_q^2)$, on peut voir que, pour tout $\beta \geqslant 1$,

$$\operatorname{tr}(\Delta L\,\Delta^T\Omega_y^{-1})^\beta = \left(\sum_{i=1}^q \sigma_i^2 \sum_{j=2}^p (\omega_{i,j} - \omega_{i,j-1})^2\right)^\beta \geqslant \sum_{i=1}^q \sigma_i^{2\beta} \sum_{j=2}^p |\omega_{i,j} - \omega_{i,j-1}|^{2\beta} \tag{2.8}$$

et la pénalisation de type *fused* apparaît clairement, mais amplifiée par β. Pour $L = I_p$, la pénalisation prend toutes les apparences d'un *ridge* (à Ω_y fixé), et donc d'un *elastic-net* une fois combinée avec $\mathrm{pen}(\Delta) = |\Delta|_1$. Comme indiqué, cette section est dédiée au contenu de l'article Okome Obiang *et al.* (2021), publié dans *ESAIM Probability and Statistics*, qui reprend ces thématiques et tente de les élargir.

Résumé

On souhaite réfléchir à un algorithme d'estimation par maximum de vraisemblance pénalisée dans un PGGM structurant dont l'objectif très général prend la forme de

$$\begin{aligned} L_n(\Omega_y, \Delta) &= -\ln\det(\Omega_y) + \operatorname{tr}(S_{n,y}\,\Omega_y) + 2\operatorname{tr}(S_{n,yx}^T\,\Delta) \\ &\quad + \operatorname{tr}(S_{n,x}\,\Delta^T\Omega_y^{-1}\Delta) + \eta\operatorname{tr}(L\,\Delta^T\Omega_y^{-1}\Delta)^\beta + \lambda\,|\Omega_y|_1^- + \mu\,|\Delta|_1 \quad (2.9) \end{aligned}$$

avec un hyperparamètre $\beta \in \mathbb{R}$ et des paramètres de régularisation $\lambda, \mu, \eta \geqslant 0$. Nous motivons par ailleurs cette écriture en expliquant en quoi elle peut être vue sous certains aspects comme le résultat de la présence d'un *a priori* gaussien généralisé sur Δ dans une hiérarchie bayésienne (cela sera formalisé dans la section suivante). Après avoir montré la convexité jointe en (Ω_y, Δ) de cet objectif lorsque $\beta \geqslant 1$ (ou $\beta = 0$), on propose une procédure d'estimation basée sur une descente de coordonnées qui alterne entre

$$\widehat{\Omega}_y = \arg\min_{\mathbb{S}_{++}^q} L_n(\Omega_y, \widehat{\Delta}) \quad \text{et} \quad \widehat{\Delta} = \arg\min_{\mathbb{R}^{q\times p}} L_n(\widehat{\Omega}_y, \Delta). \tag{2.10}$$

Plus précisément, les étapes d'optimisation sont effectuées grâce à un algorithme de type *Orthant-Wise Limited-Memory Quasi-Newton* (OWL-QN), Ω_y est estimée de façon triangulaire pour assurer sa symétrie et l'on fixe l'objectif à $+\infty$ sur $\bar{\mathbb{S}}_{++}^q$ pour imposer une solution définie positive. On montre alors, en résumé, qu'avec une forte probabilité et à condition que le modèle soit correctement régulé par le triplet (λ, μ, η),

$$\left\|\widehat{\theta} - \theta\right\|_F \lesssim \sqrt{\frac{|S_\theta|\,\ln p}{n}} \tag{2.11}$$

où $\theta = (\Omega_y, \Delta) \in \mathbb{R}^{q\times(q+p)}$ et $|S_\theta|$ est le nombre de coordonnées non-nulles de θ. Cette borne est similaire à celle de Yuan et Zhang (2014) mais également à celle de l'erreur ℓ_2 du Lasso, voir par exemple (Hastie *et al.*, 2015, Chap. 11). Un exemple sur données réelles illustre le fonctionnement de la procédure. On pourra trouver les programmes d'optimisation et de démonstration sur le GitHub `https://github.com/FredericProia/StructPGGM`.

Perspectives

L'approche est intéressante car très flexible du point de vue de la régularisation, mais elle manque encore d'un atout essentiel aux régressions en grande dimension : la possibilité d'imposer des structures de groupe. Cela ne devrait pas être insurmontable, ni en pratique ni dans la garantie théorique dans la mesure où la convexité de l'objectif est maintenue. Par contre, en dehors du cas trivial où $\beta = 0$, notre preuve de convexité repose sur la condition $\beta \geqslant 1$. Or, comme nous le reverrons dans la section suivante, il est usuel de placer un *a priori* Laplace sur le paramètre du Lasso bayésien, voir par exemple (Hastie *et al.*, 2015, Sec. 6.1), et cela correspondrait au cas $\beta = 1/2$ dans notre étude. D'autant que combiné avec la matrice des différences finies du premier ordre comme choix de L, il s'ensuivrait une pénalisation ayant les effets de la version *fused* du Lasso. En somme, il pourrait être instructif de développer une procédure valable dans cette situation particulière où la convexité de l'objectif n'est pas acquise. Par ailleurs, le domaine de

validité de la garantie théorique impose en particulier $\lambda > 0$ alors que, en pratique, il fait sens de retenir $\lambda = 0$. D'une part car q étant petit, la pénalisation des éléments de Ω_y ne paraît pas nécessaire et n'engendre d'ailleurs que peu de changement dans les résultats numériques, d'autre part et surtout car les temps de calcul en sont fortement impactés. Au vu des étapes de la preuve, cela ne paraît pas évident de premier abord mais ce cas de figure pourrait mériter une étude plus poussée. Enfin, l'optimisation de l'objectif (2.9) est sans doute atteignable par d'autres approches que celle utilisée ici : par exemple, sous l'hypothèse de convexité, peut-on annuler son gradient par un algorithme stochastique et en déduire des propriétés asymptotiques *via* les outils adaptés à ce contexte ?

A PARTIAL GRAPHICAL MODEL WITH A STRUCTURAL PRIOR ON THE DIRECT LINKS BETWEEN PREDICTORS AND RESPONSES

EUNICE OKOME OBIANG, PASCAL JÉZÉQUEL, AND FRÉDÉRIC PROÏA

Abstract. This paper is devoted to the estimation of a partial graphical model with a structural Bayesian penalization. Precisely, we are interested in the linear regression setting where the estimation is made through the direct links between potentially high-dimensional predictors and multiple responses, since it is known that Gaussian graphical models enable to exhibit direct links only, whereas coefficients in linear regressions contain both direct and indirect relations (due *e.g.* to strong correlations among the variables). A smooth penalty reflecting a generalized Gaussian Bayesian prior on the covariates is added, either enforcing patterns (like row structures) in the direct links or regulating the joint influence of predictors. We give a theoretical guarantee for our method, taking the form of an upper bound on the estimation error arising with high probability, provided that the model is suitably regularized. Empirical studies on synthetic data and a real dataset are conducted.

AMS 2020 subject classifications: Primary 62A09, 62F30; Secondary 62J05.

1. Introduction

We are interested in the recovery and estimation of direct links between high-dimensional predictors and a set of responses. Whereas the graphical models seem a natural way to go, we propose to take account of a prior knowledge on the predictors, when possible. This is typically the case when dealing with genetic markers whose joint influence may be anticipated thanks to some kind of genetic distance, or when the predictors are supposed to represent a continuous phenomenon so that consecutive covariates probably act together. In this regard, while taking up the graphical approach, we introduce some Bayesian information in a structural regularization of the estimation procedure, although the inference remains frequentist, thereby following the idea of Chiquet *et al.* [7]. This strategy also enables to affect the amount of shrinkage by playing with some hyperparametrization in the prior, while sparsity may be obtained *via* usual penalty-based patterns. Regarding the mathematical formalization of the graphical models that we will just briefly discuss in this introduction, we refer the reader to the very complete handbook recently edited by Maathuis *et al.* [16]. We also refer the reader to the book of Hastie *et al.* [11] and to the one of Giraud [10], both related to the standard high-dimensional statistical methods. Before introducing the model and the organization of this work, let us describe the notation used throughout the paper.

1.1. **Notation.** For any matrix A, $|A|_* = \|\mathrm{vec}(A)\|_*$ is the elementwise ℓ_* norm of A and $|A|_*^-$ is $|A|_*$ deprived of the diagonal terms of A. We also note $\|A\|_F = |A|_2$ the Frobenius norm of A and $\|A\|_2$ the spectral norm of A. The Frobenius inner product between any matrices A and B of same dimensions is $\langle\langle A, B\rangle\rangle = \langle \mathrm{vec}(A), \mathrm{vec}(B)\rangle = \mathrm{tr}(A^t B)$ whereas

Key words and phrases. High-dimensional linear regression, Partial graphical model, Structural penalization, Sparsity, Convex optimization.

$\langle u, v\rangle = u^t\, v$ is the inner product of the Euclidean real space. For any vector u, $|u|_0$ is the number of non-zero values in u. For a matrix A, $[A]_C$ is to be understood as the matrix A whose elements outside of the set of coordinates C are set to zero and $\mathrm{vec}(A)$ is the vectorization of A into a column vector. The eigenvalues of a square matrix A of size d with spectrum $\mathrm{sp}(A)$ are $\lambda_i(A)$ taken in decreasing order (from $\lambda_1(A) = \lambda_{\max}(A)$ to $\lambda_d(A) = \lambda_{\min}(A)$). The cones of symmetric positive semi-definite and definite matrices of dimension d are $\mathbb{S}_+^d$ and $\mathbb{S}_{++}^d$ respectively.

1.2. **The partial graphical model.** In the classic Gaussian graphical model (GGM) setting, we aim at estimating the precision matrix $\Omega = \Sigma^{-1}$ of jointly normally distributed random vectors $Y \in \mathbb{R}^q$ and $X \in \mathbb{R}^p$ with zero mean and covariance Σ. The point is that it induces a graphical structure among the variables and the support of Ω is closely related to the conditional interdependences between them. Let us consider, now and in all the study, the sample covariances of n independent observations (Y_i, X_i), denoted by

$$S_{yy}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} Y_i\, Y_i^t, \quad S_{yx}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} Y_i\, X_i^t \quad \text{and} \quad S_{xx}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i\, X_i^t. \tag{1.1}$$

Maximizing the penalized likelihood of a GGM boils down to finding $\Omega \in \mathbb{S}_{++}^{p+q}$ that minimizes the convex objective

$$L_n(\Omega) = -\ln\det(\Omega) + \langle\!\langle S^{(n)}, \Omega\rangle\!\rangle + \lambda\,\mathrm{pen}(\Omega) \tag{1.2}$$

where $S^{(n)}$ is the full sample covariance built from the blocks (1.1). The penalty function $\mathrm{pen}(\Omega)$ is usually $|\Omega|_1$ or even $|\Omega|_1^-$. Efficient algorithms exist to get solutions for (1.2), see *e.g.* Banerjee *et al.* [2], Yuan and Lin [28], Lu [15] or the graphical Lasso of Friedman *et al.* [9]. The reader may also look at the theoretical guarantees of Ravikumar *et al.* [21]. However, thinking at X_i as a predictor of size p associated with a response Y_i of size q, the partial Gaussian graphical model (PGGM), developped *e.g.* by Sohn and Kim [26] or Yuan and Zhang [29], appears as a powerful tool to exhibit direct relationships between the predictors and the responses. To understand this, consider the decomposition into blocks

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{yx}^t & \Omega_{xx} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx}^t & \Sigma_{xx} \end{pmatrix}$$

where $\Omega_{yy} \in \mathbb{S}_{++}^q$, $\Omega_{yx} \in \mathbb{R}^{q\times p}$ and $\Omega_{xx} \in \mathbb{S}_{++}^p$ and where the same goes for Σ_{xx}, Σ_{yx} and Σ_{xx}. The precision matrix $\Omega = \Sigma^{-1}$ satisfies, by blockwise inversion,

$$\Omega_{yy}^{-1} = \Sigma_{yy} - \Sigma_{yx}\,\Sigma_{xx}^{-1}\,\Sigma_{yx}^t \quad \text{and} \quad \Omega_{yx} = -(\Sigma_{yy} - \Sigma_{yx}\,\Sigma_{xx}^{-1}\,\Sigma_{yx}^t)^{-1}\,\Sigma_{yx}\,\Sigma_{xx}^{-1}. \tag{1.3}$$

The conditional distribution peculiar to Gaussian vectors

$$Y_i \,|\, X_i \sim \mathcal{N}(-\Omega_{yy}^{-1}\,\Omega_{yx}\,X_i,\, \Omega_{yy}^{-1})$$

gives a new light on the multiple-output regression $Y_i = B^t\, X_i + E_i$ with Gaussian noise $E_i \sim \mathcal{N}(0, R)$, through the reparametrization $B = -\Omega_{yx}^t\,\Omega_{yy}^{-1}$ and $R = \Omega_{yy}^{-1}$. Whereas B contains direct and indirect links between the predictors and the responses (due *e.g.* to strong correlations among the variables), Ω_{yx} only contains direct links, as it is shown by the graphical models theory. In other words, the direct links are closely related to the concept of partial correlations between X and Y (see Meinshausen and Bülmann [17] or Peng *et al.* [19], for the univariate case). For example, the direct link between predictor k and response ℓ may be evaluated through the partial correlation $\mathbb{C}\mathrm{orr}(Y_\ell, X_k \,|\, Y_{\neq\ell}, X_{\neq k})$ contained, apart from a

multiplicative coefficient, in the ℓ-th row and k-th column of Ω_{yx} (see *e.g.* Cor. A.6 in [10]) with the particularly interesting consequence that the support of Ω_{yx} is sufficient to identify direct relationships between X and Y. Hence, in the partial setting, the objective reduces to the estimation of the direct links Ω_{yx} together with the conditional precision matrix of the responses Ω_{yy}. Maximizing the penalized conditional log-likelihood of the model now comes down to minimizing the new convex objective

$$\begin{aligned} L_n(\Omega_{yy}, \Omega_{yx}) \;=\; & -\ln\det(\Omega_{yy}) + \langle\!\langle S_{yy}^{(n)}, \Omega_{yy} \rangle\!\rangle + 2\,\langle\!\langle S_{yx}^{(n)}, \Omega_{yx} \rangle\!\rangle \\ & + \langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx} \rangle\!\rangle + \lambda\,\mathrm{pen}(\Omega_{yy}) + \mu\,\mathrm{pen}(\Omega_{yx}) \end{aligned} \tag{1.4}$$

over $(\Omega_{yy}, \Omega_{yx}) \in \mathbb{S}_{++}^q \times \mathbb{R}^{q\times p}$ for some usual penalty functions. It is worth noting that $\mathrm{pen}(\Omega_{yx})$ often plays a crucial role in modern statistics dealing with high-dimensional predictors (and the natural choice is $|\Omega_{yx}|_1$ to get sparsity) while we may choose $\lambda = 0$, because the number of responses is generally small. In the seminal papers [26] and [29], the authors consider $|\Omega_{yy}|_1$ and $|\Omega_{yy}|_1^-$ for $\mathrm{pen}(\Omega_{yy})$, respectively. Yuan and Zhang [29] also point out that no estimation of Ω_{xx} is needed anymore. In a graphical model, the estimation of Ω_{yx} and Ω_{yy} depends on the accuracy of the estimation of Ω which, in turn, is strongly affected by the one of Ω_{xx}, especially in a high-dimensional setting. The partial model overrides this issue, the focus is on Ω_{yx} and Ω_{yy} while Ω_{xx} has disappeared from the objective function (1.4). The latter is obtained either by considering the multiple-output Gaussian regression scheme, or, as it is done in [29], by eliminating Ω_{xx} thanks to a first optimization step in (1.2). In this paper, we will consider the penalties

$$\mathrm{pen}(\Omega_{yy}) = |\Omega_{yy}|_1^- \quad \text{and} \quad \mathrm{pen}(\Omega_{yx}) = |\Omega_{yx}|_1 \tag{1.5}$$

which correspond to the PGGM (Gm) of [29]. The Spring (Spr) of [7] can also be seen as a PGGM but with no penalty on Ω_{yy} (replaced with an additional structuring one on Ω_{yx}, we will come back to this point thereafter), so for Spr we may consider $\lambda = 0$. The generalized procedure (GenGm) at the heart of the study relies on a combination between these two approaches. We will see in due time that we keep both the penalties of Gm and the structuring one of Spr on Ω_{yx}. Finally, the intermediate solution consisting in estimating Ω_{yy} and B through the conditional distribution $Y_i \,|\, X_i \sim \mathcal{N}(B^t\,X_i,\, \Omega_{yy}^{-1})$ with penalizations both on B and Ω_{yy}, presented and analyzed by Rothman *et al.* [23] and by Lee and Liu [14], is better known as a multivariate regression with covariance estimation (MRCE). However, it has been shown that the objective function suffers from a lack of convexity and that the optimization procedure may be debatable, in addition to the less convenient setup for statistical interpretation (B contains both direct and indirect influences) compared to PGGM. Without claiming to be exhaustive, let us conclude this quick introduction by citing some related works, like the structural generalization of the Elastic-Net of Slawski *et al.* [25], the Dantzig approach of Cai *et al.* [6] put in practice on genomic data [5], the greedy research of the non-zero pattern in Ω of Johnson *et al.* [13], the approach of Fan *et al.* [8] using a non-convex SCAD penalty to reduce the bias of the Lasso in the estimation of Ω, the eQTL data analysis of Yin and Li [27] which makes use of a sparse conditional GGM, and so on. All the references inside will complete this concise list.

1.3. **Organization of the paper.** To sum up, we have two goals in this paper:

(1) Give some theoretical guarantees to the (slightly modified) model introduced in Chiquet *et al.* [7].

(2) Generalize the result of Yuan and Zhang [29] to the case where a structural penalization is added in the estimation step.

In Section 2, we introduce the model, consisting in putting a generalized Gaussian prior on the direct links before the procedure of estimation of Ω_{yy} and Ω_{yx}, and we detail the new convex objective. Then we provide some error bounds for our estimates, useful as theoretical guarantees of performance. Section 3 is devoted to empirical considerations. We explain how we deal with the minimization of the new objective and we test the method on simulations first, and next on a real dataset (a Canadian average annual weather cycle, see *e.g.* [20]). After a short conclusion in Section 4, we finally prove our results in Section 5. The numerous constants appearing in the results and the proofs are gathered in the Appendix, for the sake of readability.

2. A generalized Gaussian prior on the direct links

We use the definition given in formulas (1)-(2) of [18] for the so-called d-dimensional multivariate generalized Gaussian $\mathcal{GN}(0, 1, V, \beta)$ distribution with mean 0, scale 1, scatter parameter $V \in \mathbb{S}_{++}^d$ and shape parameter $\beta > 0$. According to the authors, the density takes the form of

$$\forall\, z \in \mathbb{R}^d, \quad f_{V,\beta}(z) = \frac{\beta\,\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}}\,\Gamma(\frac{d}{2\beta})\,2^{\frac{d}{2\beta}}\sqrt{\det(V)}} \exp\left(-\frac{\langle z, V^{-1}z\rangle^{\beta}}{2}\right)$$

where Γ is the Euler Gamma function.

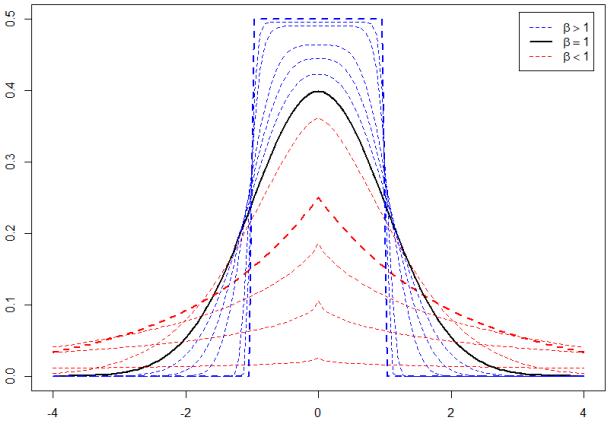


FIGURE 1. Marginal shape of the generalized Gaussian distribution ($d = 1$ and $V = 1$) for some $\beta < 1$ (dotted red), $\beta = 1$ (black) and some $\beta > 1$ (dotted blue). The noteworthy cases $\beta = 1/2$ (Laplace), $\beta = 1$ (Gaussian) and $\beta = +\infty$ (uniform) are highlighted.

We clearly recognize the Gaussian $\mathcal{N}(0, V)$ setting for $\beta = 1$. Moreover, for $\beta = 1/2$, it can be seen as a multivariate Laplace distribution whereas it is known to converge to some uniform distribution as $\beta \to +\infty$. The marginal shapes ($d = 1$ and $V = 1$) of the distribution

are represented on Figure 1, depending on whether $\beta < 1$, $\beta = 1$ or $\beta > 1$. Our results hold for all $\beta \geqslant 1$ but, as will be explained in due course, we shall not theoretically deviate too much from the Gaussianity in the prior (even if we will allow ourselves some exceptions in the practical works). The usual Bayesian approach for multiple-output Gaussian regression having B as matrix of coefficients and R as noise variance consists in a conjugate prior $\mathrm{vec}(B) \sim \mathcal{N}(b, R \otimes L^{-1})$ for some information matrix $L \in \mathbb{S}_{++}^p$ and a centering value b (see *e.g.* Sec. 2.8.5 of [22]). In the PGGM reformulation, we have $R = \Omega_{yy}^{-1}$ and $B = -\Omega_{yx}^t\,\Omega_{yy}^{-1}$ as explained in Section 1, and of course we shall choose $b = 0$ to meet our purposes. Thus,

$$\mathrm{vec}(\Omega_{yx}^t) = -(\Omega_{yy} \otimes I_p)\,\mathrm{vec}(B) \;\sim\; \mathcal{N}(0, \Omega_{yy} \otimes L^{-1})$$

is a natural prior for the direct links (this is in particular the choice of the authors of [7]). Following the same logic, let us choose $\Omega_{yy} \otimes L^{-1}$ for scatter parameter and suppose that

$$\mathrm{vec}(\Omega_{yx}^t) \;\sim\; \mathcal{GN}(0, 1, \Omega_{yy} \otimes L^{-1}, \beta). \tag{2.1}$$

In this way, we can play on the intensity of the constraint we want to bring on Ω_{yx}, from a non-informative prior to quasi-boundedness through Laplace and Gaussian distributions. This prior entails an additional smooth term acting as a structural penalization in the objective (1.4) that becomes

$$\begin{aligned} L_n(\Omega_{yy}, \Omega_{yx}) \;&=\; -\ln\det(\Omega_{yy}) + \langle\!\langle S_{yy}^{(n)}, \Omega_{yy} \rangle\!\rangle + 2\,\langle\!\langle S_{yx}^{(n)}, \Omega_{yx} \rangle\!\rangle \\ &\quad + \langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx} \rangle\!\rangle + \eta\,\langle\!\langle L, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx} \rangle\!\rangle^\beta + \lambda\,|\Omega_{yy}|_1^- + \mu\,|\Omega_{yx}|_1 \end{aligned} \tag{2.2}$$

with three regularization parameters (λ, μ, η). The smooth penalization lends weight to the prior on Ω_{yx} and thereby plays on the extent of shrinkage and structuring through β, whereas $|\Omega_{yx}|_1$ and $|\Omega_{yy}|_1^-$ are designed to induce sparsity. One can note that this is closely related to the log-likelihood of a hierarchical model of the form

$$\left\{ \begin{array}{l} Y_i \,|\, X_i, \Omega_{yx} \;\sim\; \mathcal{N}(-\Omega_{yy}^{-1}\,\Omega_{yx}\,X_i,\, \Omega_{yy}^{-1}) \\ \mathrm{vec}(\Omega_{yx}^t) \;\sim\; \mathcal{GN}(0, 1, \Omega_{yy} \otimes L^{-1}, \beta) \end{array} \right.$$

where the emphasis is on Ω_{yx} in the prior and Ω_{yy} remains a fixed parameter, although it is important to see that, in this work, the estimation step does not rely on a posterior distribution. The following proposition is related to the existence of a global minimum for our objective (2.2) with respect to $(\Omega_{yy}, \Omega_{yx})$ as soon as $\beta \geqslant 1$.

Proposition 2.1. *Assume that $\beta \geqslant 1$. Then, $L_n(\Omega_{yy}, \Omega_{yx})$ defined in* (2.2) *is jointly convex with respect to $(\Omega_{yy}, \Omega_{yx})$.*

Proof. See Section 5.2. □

Now and throughout the rest of the paper, denote by $\theta = (\Omega_{yy}, \Omega_{yx}) \in \Theta = \mathbb{S}_{++}^q \times \mathbb{R}^{q \times p}$ the $(q \times (q+p))$-matrix of parameters of the model, with true value $\theta^* = (\Omega_{yy}^*, \Omega_{yx}^*)$. As it is usually done in studies implying sparsity, we will also consider S of cardinality $|S|$, the true active set of θ^* defined as $S = \{(i,j),\, \theta_{i,j}^* \neq 0\}$, and its complement $\bar{S}$. Our results also depends on some basic assumptions related to the true covariances of the Gaussian observations, and we will assume that the following holds.

$$(\mathrm{H}_1) \qquad \Sigma_{xx}^* \in \mathbb{S}_{++}^p, \quad \Omega_{yy}^* \in \mathbb{S}_{++}^q, \quad B \neq 0 \text{ (that is, } \Omega_{yx}^* \neq 0) \quad \text{and} \quad \Omega_{yx}^*\,L\,\Omega_{yx}^{*\,t} \in \mathbb{S}_{++}^q.$$

This is a natural hypothesis in our framework, in particular we suppose that there is at least a link between X and Y.

Remark 2.1 (Null model). Even if it is of less interest, our study does not exclude the case where $\Omega^*_{yx} = 0$. Indeed, we might as well consider that $\Omega^*_{yx} = 0$ and get the same results, but some constants should be refined. On the other hand, $\Sigma^*_{xx} \in \mathbb{S}^p_{++}$ and $\Omega^*_{yy} \in \mathbb{S}^q_{++}$ are crucial.

Under (H_1), the random matrices

$$A_n = (S^{(n)}_{yy} - \Sigma^*_{yy}) - \Omega^{*\,-1}_{yy}\,\Omega^*_{yx}\,(S^{(n)}_{xx} - \Sigma^*_{xx})\,\Omega^{*\,t}_{yx}\,\Omega^{*\,-1}_{yy} \quad \text{with} \quad h_a = |A_n|_\infty \tag{2.3}$$

and

$$B_n = 2\,((S^{(n)}_{yx} - \Sigma^*_{yx}) + \Omega^{*\,-1}_{yy}\,\Omega^*_{yx}\,(S^{(n)}_{xx} - \Sigma^*_{xx})) \quad \text{with} \quad h_b = |B_n|_\infty \tag{2.4}$$

are going to play a fundamental role, especially h_a and h_b. Let us now provide some theoretical guarantees for the estimation of θ in our model, provided that the regularization parameters are located in a particular area $(\lambda, \mu, \eta) \in \Lambda$. Consider the penalized likelihood $\ell_{\lambda,\mu,\eta}(\theta)$ given in (2.2), and estimate θ by the global minimum

$$\widehat{\theta} = \arg\min_{\Theta}\, \ell_{\lambda,\mu,\eta}(\theta) \tag{2.5}$$

obtained for $\beta \geqslant 1$. To facilitate reading, we postpone the precise definition of the numerous constants to the Appendix. We recall that p is the number of predictors, q is the number of responses and $|S|$ is the size of the true active set.

Theorem 2.1. *Fix $d_\lambda > c_\lambda > 1$, $d_\mu > c_\mu > 1$, $e_\lambda > 0$ and $e_\mu > 0$, and assume that the regularization parameters satisfy $(\lambda, \mu, \eta) \in \Lambda = [c_\lambda\, h_a,\, d_\lambda\, h_a] \times [c_\mu\, h_b,\, d_\mu\, h_b] \times [0,\, \overline{\eta}]$, where*

$$\overline{\eta} = \frac{\min\left\{ \frac{(c_\lambda - 1)\,\lambda}{c_\lambda\,\ell_a},\ \frac{(c_\mu - 1)\,\mu}{c_\mu\,\ell_b},\ \frac{e_\lambda\, h_a}{\ell_a},\ \frac{e_\mu\, h_b}{\ell_b} \right\}}{\beta\, s_L^{\beta - 1}}$$

for some non-random constants s_L, ℓ_a and ℓ_b defined in (A.2) *and* (A.3)*, and the random constants h_a and h_b given above. Then, under* (H_1)*, there exists absolute constants $b_1 > 0$ and $b_2 > 0$ such that, for any $0 < b_3 < 1$ and as soon as $n > n_0$, with probability no less that $1 - \mathrm{e}^{-b_2 n} - b_3$, the estimator* (2.5) *satisfies*

$$\|\widehat{\theta} - \theta^*\|_F \ \leqslant\ \frac{16\, m^*\, c_{\lambda,\mu}\sqrt{|S|}}{\gamma_{r,\eta,\beta,p}} \sqrt{\frac{\ln(10(p+q)^2) - \ln(b_3)}{n}}$$

where $\gamma_{r,\eta,\beta,p}$, $c_{\lambda,\mu}$ and m^* are technical constants defined in (A.7), (A.8) *and* (A.9)*, respectively, and where the minimal number of observations is given by*

$$n_0 \ = \ \max\left\{ \frac{(\ln(10(p+q)^2) - \ln(b_3))\, c^2_{\lambda,\mu}\, |S|\, (16\, m^*)^2}{r^{*\,2}\, \gamma^2_{r,\eta,\beta,p}}, \right.$$

$$\left. b_1\,(q + \lceil s_\alpha \rceil \ln(p+q)), \ln(10(p+q)^2) - \ln(b_3) \right\} \tag{2.6}$$

with s_α defined in (A.5) *and r^* in* (A.6)*.*

Proof. See Section 5.3. □

Among all these constants, we can note that s_L, ℓ_a, ℓ_b, h_a and h_b are useful to properly describe and restrict Λ, the domain of validity of (λ, μ, η) for the theorem to hold. Once Λ is fixed, the other constants take part in the upper bound of the estimation error. However, as it stands, the theorem is very difficult to interpret. The next two remarks seem essential to have an overview of the orders of magnitude involved for the number of observations, for p and q, for the estimation error and for the regularization parameters.

Remark 2.2 (Validity band). Of course the degree of sparsity $|S|$ is crucial in the estimation error, but it also plays an indirect role in the probability associated with the theorem and in the numerous constants. In virtue of Lemma 5.12, we can hope that λ and μ have a wide validity band, by playing on c_λ, c_μ, d_λ and d_μ. In turn, η also has a non-negligible area of validity, provided of course that ℓ_a, ℓ_b and s_L, all depending on combinations between Ω^*_{yx}, Ω^{*-1}_{yy} and L, are small enough. Accordingly, it would be to our advantage if L was both sparse and not chosen with too large elements. As it always appears together with η, we may as well take a normalized version of L (*e.g.* $|L|_\infty \leqslant 1$).

Remark 2.3 (Order of magnitude). Even if the result holds for any $\beta \geqslant 1$, the terms $\propto p^{\beta-1}$ appearing in some upper bounds of the proof clearly argue in favor of a moderate choice $\beta \in [1, 1+\epsilon]$ for a small $\epsilon > 0$, depending on p. In other words, we cannot deviate too much from the Gaussianity in the prior on the direct links. For example in a very high-dimensional setting ($p \sim 10^7$), choosing $\epsilon = 0.1$ leads to $p^{\beta-1} \approx 5$ whereas we may try larger values of ϵ for the more common high-dimensional settings $p \sim 10^3$ or $p \sim 10^4$. By contrast, we can see that n_0 must (at least) grow like q for the theorem to hold, so high-dimensional responses are excluded. However in multiple-output regressions, even when p is extremely large, q generally remains small. According to all these considerations, we may roughly say that, in a high-dimensional setting with respect to p,

$$\|\widehat{\theta} - \theta^*\|_F \lesssim \sqrt{\frac{|S| \ln p}{n}}$$

with a large probability, under a suitable regularization of the model. We recognize the usual terms appearing in the error bounds of regressions with high-dimensional covariates, like the ℓ_2 error of the Lasso (see *e.g.* Chap. 11 of [11]). This is the same bound as in [29], but our additional structural penalty restricts Λ.

3. Simulations and real dataset

The minimization problem (2.5) is solved using a coordinate descent procedure, alternating between the computations of

$$\widehat{\Omega}_{yy} = \arg\min_{\mathbb{S}^q_{++}} \ell_{\lambda,\mu,\eta}(\Omega_{yy}, \widehat{\Omega}_{yx}) \quad \text{and} \quad \widehat{\Omega}_{yx} = \arg\min_{\mathbb{R}^{q\times p}} \ell_{\lambda,\mu,\eta}(\widehat{\Omega}_{yy}, \Omega_{yx}).$$

Each step is done by an Orthant-Wise Limited-Memory Quasi-Newton (OWL-QN) algorithm (see *e.g.* [1]). The first subproblem is performed through half-vectorization (vech) to ensure symmetry and we set the objective to $+\infty$ on $\bar{\mathbb{S}}^q_{++}$ to ensure positive definiteness of the solution. The coordinate descent is stopped when

$$\|\widehat{\Omega}^{(t)}_{yy} - \widehat{\Omega}^{(t-1)}_{yy}\|_2 \leqslant \epsilon \max(1, \|\widehat{\Omega}^{(t-1)}_{yy}\|_2) \quad \text{and} \quad \|\widehat{\Omega}^{(t)}_{yx} - \widehat{\Omega}^{(t-1)}_{yx}\|_2 \leqslant \epsilon \max(1, \|\widehat{\Omega}^{(t-1)}_{yx}\|_2)$$

following two consecutive iterations $t-1$ and t, where $\epsilon > 0$ is a small threshold depending on the desired precision. We are now going to try our method on synthetic data first, and then on a real dataset. We will pay attention to the role played by β, in particular we will see that it can be useful as well as counterproductive, depending on the situations.

3.1. **Simulations.** For each scenario, we first generate i.i.d. standard Gaussian vectors $X_i \in \mathbb{R}^p$, then $Y_i \in \mathbb{R}^q$ is simulated according to the setting and we estimate Ω_{yy} and Ω_{yx}. From the relations detailed in Section 1, we recall that $Y_i = B^t X_i + E_i$ with $E_i \sim \mathcal{N}(0, R)$ is an equivalent formulation, provided that $B = -\Omega_{yx}^t \, \Omega_{yy}^{-1}$ and $R = \Omega_{yy}^{-1}$. In a compact form, we may also write

$$Y = XB + E \quad \text{or} \quad \operatorname{vec}(Y) = (I_q \otimes X) \operatorname{vec}(B) + \operatorname{vec}(E)$$

where the i-th row of Y is ${Y_i}^t$ and the i-th row of X is X_i^t. Thus, we can estimate B using the Lasso (Las) and the Group-Lasso (GLas) in the vectorized form, to provide a basis for comparison between our method and the usual penalized methods. The Lasso penalty is obviously $\|\operatorname{vec}(B)\|_1$ to promote coordinate sparsity while, for the Group-Lasso, we use the penalty $\|B_1\|_2 + \ldots + \|B_p\|_2$ where B_i is the i-th row of B, to promote row sparsity and exclude altogether some predictors from the model. We also implement some variants of our generalized graphical model (GenGm). The case where $\Omega_{yy} = R^{-1}$ is known and does not need to be estimated is the Oracle (Or) and the case where $\eta = 0$ so that β has no influence is the classic PGGM (Gm). The case where $\lambda = 0$ and $\beta = 1$ is called the Spring (Spr) by the authors of [7]. We will focus on structured scenarios. With no structure in Ω_{yx}, there is no reason why our method should outperform the usual PGGM. In a completely random setting, we have observed that all PGGM procedures perform identically. In fact, a slight gain can be obtained compared to Spr and Gm simply due to the flexibility induced by the additional parameter (Spr and Gm are particular cases of GenGm). However, that clearly cannot counterbalance the extended computational times, and GenGm should not be used for such situations. The calibration of the regularization parameters is made using a cross-validation on a training set of size $n_t = 150$ and the accuracy is evaluated thanks to the mean squared prediction error (MSPE) on a validation set of size $n_v = 1000$,

$$\text{MSPE} = \frac{\left\| Y + X \, \widehat{\Omega}_{yx}^t \, \widehat{\Omega}_{yy}^{-1} \right\|_F^2}{q \, n_v}. \tag{3.1}$$

Due to the large amount of treatments, the grids for cross-validation are not very sharp here but they will be carefully refined for the real datasets of the next section. The covariance between the outputs is $R = (r^{|i-j|})_{1 \leqslant i,\, j \leqslant q}$ for $r = \frac{1}{2}$ and we work with $p = 100$. Each scenario is repeated $N = 500$ times and GenGm is evaluated with numerous values of β, from 0.25 to 2 with a step of 0.25. The results of the following scenarios are summarized on Figures 2, 3 and 4 below, respectively.

- → Scenario 1 ($q = 1$). We draw $\omega_i = \pm\frac{1}{2}$ for $i = 1, \ldots, 10$ and we fill 10 randomly selected sections of size 3 in Ω_{yx} with ω_i. The remaining part of Ω_{yx} is 0.
- → Scenario 2 ($q = 2$). We draw $\omega = \pm\frac{1}{2}$ and one randomly selected row of Ω_{yx} is filled with ω while the other is identically 0.
- → Scenario 3 ($q = 3$). We draw $\omega_i = \pm\frac{1}{2}$ and we fill a randomly selected section of size 30 on the i-th row of Ω_{yx} with ω_i, for $i = 1, 2, 3$. The remaining part of Ω_{yx} is 0.

The row structure is promoted by a normalized first finite difference operator

$$L = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix} \tag{3.2}$$

which, through $\Omega_{yx}\, L\, \Omega_{yx}^t$, tends to penalize the difference between two consecutive values on a same row (as does Fused-Lasso with ℓ_1 penalty). Yet, the Fused-Lasso is not a suitable alternative to GLas and Las in this precise context because $B = -\Omega_{yx}^t\, \Omega_{yy}^{-1}$ is not supposed to have a row structure even if Ω_{yx} has one. For this choice of L, one can note that, in the particular case where $R = \mathrm{diag}(\sigma_1^2, \dots, \sigma_q^2)$,

$$\langle\!\langle L, \Omega_{yx}^t\, \Omega_{yy}^{-1}\, \Omega_{yx} \rangle\!\rangle^\beta = \left(\sum_{i=1}^{q} \sigma_i^2 \sum_{j=2}^{p} (\omega_{i,j} - \omega_{i,j-1})^2 \right)^\beta \geqslant \sum_{i=1}^{q} \sigma_i^{2\beta} \sum_{j=2}^{p} |\omega_{i,j} - \omega_{i,j-1}|^{2\beta}$$

where $\omega_{i,j}$ is the (i,j)-th element of Ω_{yx}, so we may fairly expect that $\beta \geqslant 1$ is going to strengthen the smoothness of the estimation and to enforce all the more the structuring.

Remark 3.1 (Validity of the hypotheses). We could as well add a small diagonal element in the matrix L defined above, positive semi-definite but not invertible. The resulting effect would be a negligible ridge-like penalization on the elements of Ω_{yx}. This is not required for the estimation procedure but useful for Theorem 2.1 to hold (see *e.g.* (H_1)). Likewise, it seemed interesting to test some settings with $\beta < 1$ even if the theory developped in the paper does not give any guarantee for them, as a basis for comparison.

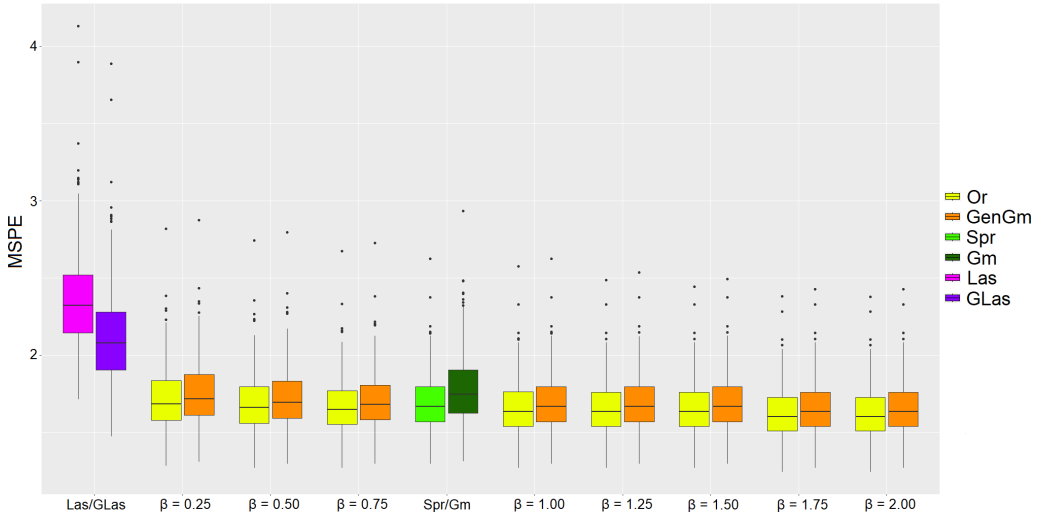


FIGURE 2. Mean squared prediction error for $N = 500$ repetitions of the weakly structured Scenario 1.

First of all, one can observe that Las and GLas are left behind in all our simulations. This is not surprising since the covariance between the outputs cannot be recovered with the

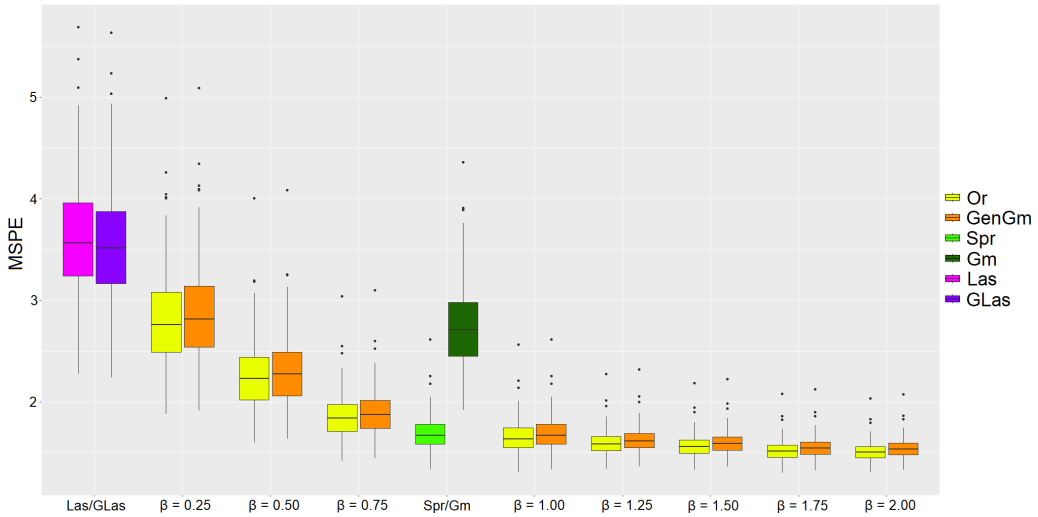


FIGURE 3. Mean squared prediction error for $N = 500$ repetitions of the strongly structured Scenario 2.

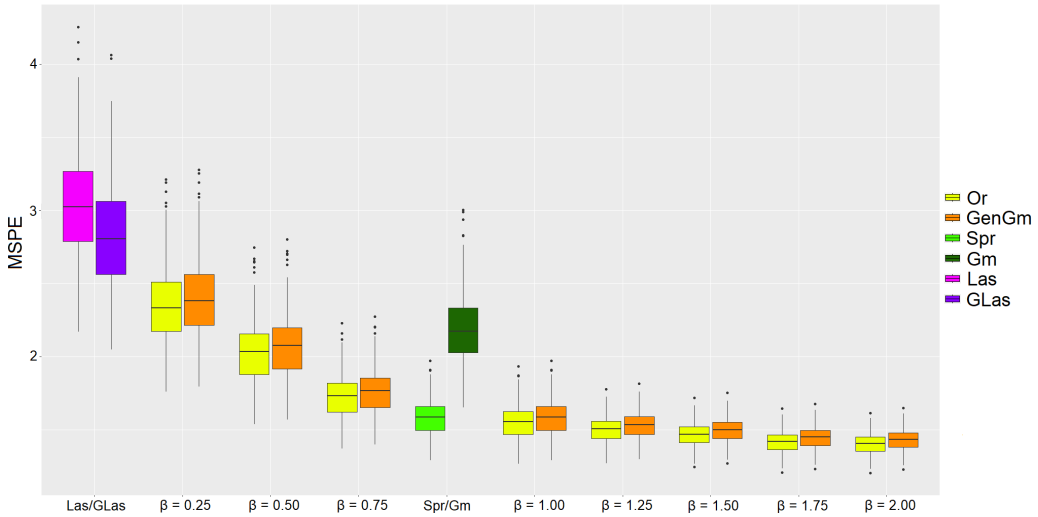


FIGURE 4. Mean squared prediction error for $N = 500$ repetitions of the strongly structured Scenario 3.

standard Lasso, at least for $q \geqslant 2$. Generally, GLas remains more robust compared to Las, probably due to the high level of sparsity in Ω_{yx} approximately passed to B (provided that the covariances in R are small enough), and exploited by the grouping effect. In the weakly structured setting (Scenario 1), we also observe that, as expected, all PGGM procedures perform almost identically, with obviously an advantage for Or (although small, illustrating the accuracy of the estimation). In the strongly structured settings (Scenarios 2 and 3), Gm gives results below the expected level, because it is not designed to promote such layouts. On the contrary, thanks to this choice of L showing here great efficiency, GenGm and Spr are doing pretty well. Note that, in this context, GenGm with $\beta = 1$ is almost the same as Spr since, q being small, λ does not play a crucial role. However, some empirical facts draw our

attention: the prediction error decreases with β to some extent, but the most interesting fact seems to be the simultaneous decrease of its variance. It is likely that the increasing pressure exerted by β on the estimation procedure leads to a higher homogeneity in the numerical results, despite the repetitions of random experiments under random settings. In other words, the structuring seems to be strengthened and we also observe that the convergence of the algorithm is faster, which logically follows from the latter remarks (especially clear when we compare $\beta = 0.25$ and $\beta = 2$). On the other hand, for the opposite reason, we notice that the predictions are hardly better than Gm (even worse in some cases), both on average and in terms of variability, for $\beta < 1$, and these simulations tend to undermine such values of the hyperparameter. On the whole, GenGm with $\beta > 1$ might be a sound approach for practitioners who place a high priority on structuring the estimations, even if Remark 3.2 below should probably temper this statement. To conclude, let us consider the strongly structured scenarios with $L = I_p$ (without structuring) in the Oracle setting with $\beta = 2$, and let us compare the results with those of Figures 3 and 4, obtained with the correct version of L given in (3.2). The results are displayed on Figure 5 where we can see that the benefit of structuring is manifest. Unsurprisingly, the results without structuring are close to those of Gm since $L = I_p$ only strengthens the shrinkage effect with ridge-like additional penalties.

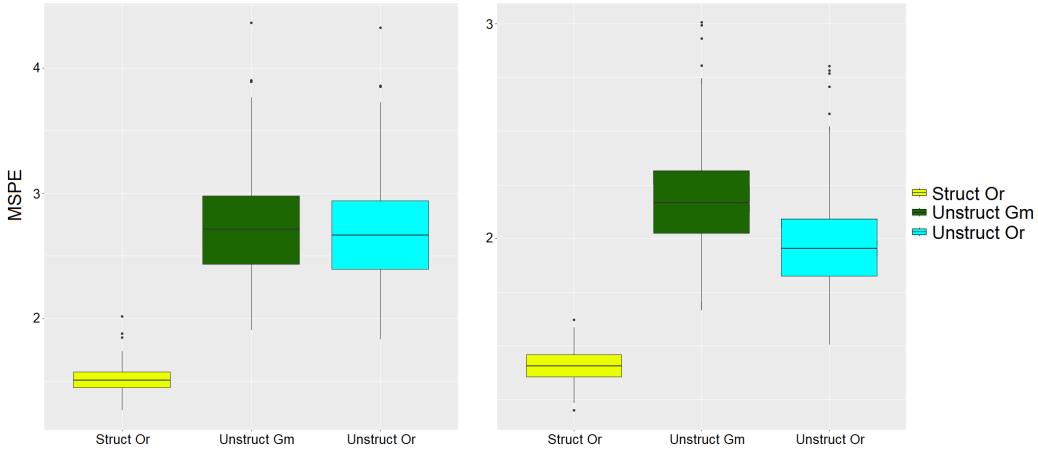


FIGURE 5. Mean squared prediction error for $N = 500$ repetitions of the strongly structured Scenario 2 (left) and Scenario 3 (right) for Or, Gm and the unstructured Or ($L = I_p$), with $\beta = 2$.

Remark 3.2 (Computational time). To estimate $(\Omega_{yy}, \Omega_{yx})$ in the model Spr, the authors of [7] use a very judicious and efficient method relying, in each step of the coordinate descent procedure, on a direct computation of the estimation of Ω_{yy} together with an Elastic-Net estimation of Ω_{yx}. This is possible for $\lambda = 0$ and $\beta = 1$, but unfortunately cannot be implemented in the general setting. As a result, computational times remain an issue that should be paid attention to.

Remark 3.3 (Oracle-type errors). The mean value of the estimation errors $\|\widehat{\Omega}_{yx} - \Omega_{yx}\|_F^2$ leads to the same kind of observations for the models being compared in the simulations. But the minimal prediction error does not always coincide with an optimal support recovery due to the shrinkage effect on the estimation of Ω_{yx}, when the coefficients or the covariates

are not very contrasting. The so-called F-score is given by

$$F = \frac{2\, p_r\, r_e}{p_r + r_e} \quad \text{where} \quad p_r = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad \text{and} \quad r_e = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

are the *precision* and the *recall*, respectively, and where T/F and P/N stand for true/false and positive/negative. In the strongly structured scenarios, F is generally located between 0.60 and 0.65, and a deeper analysis shows that a proportion of more than 0.99 of true non-zero values are recovered (that is, the part of the true active set S related to Ω_{yx}). If the models are not calibrated to reach the best prediction error but the best F-score, F regularly exceeds 0.90, at least for the structured procedures.

Nevertheless, Scenarios 2 and 3 are very strongly structured, more than one would expect from an unknown underlying generating process, and the real dataset of the next section is going to highlight the fact that the improvement may be hardly noticeable with respect to β. But we will see that β can still be useful for variable selection.

3.2. **A real dataset.** The dataset available as `CanadianWeather` in the `R` package `fda` contains daily temperature and precipitation at 35 different locations in Canada, averaged over annual reports starting in 1960 and ending in 1994 (see *e.g.* [20]). We intend to look at the direct links between the minimal and maximal rainfall (on the $\log_{10}$ scale) and the temperature pattern in the 35 weather stations, so as to identify the times of the year that have a strong effect on rainfall (positive as well as negative). In this context, $n = 35$, $q = 2$ and $p = 365$. Figure 6 shows temperature and log-precipitation measured over a year in Montreal, chosen as an example, together with the empirical distribution of the minimal and maximal log-precipitation for the 35 weather stations. We can note that, since the data are averaged over numerous years, outliers are unlikely even for the extremes (min and max).

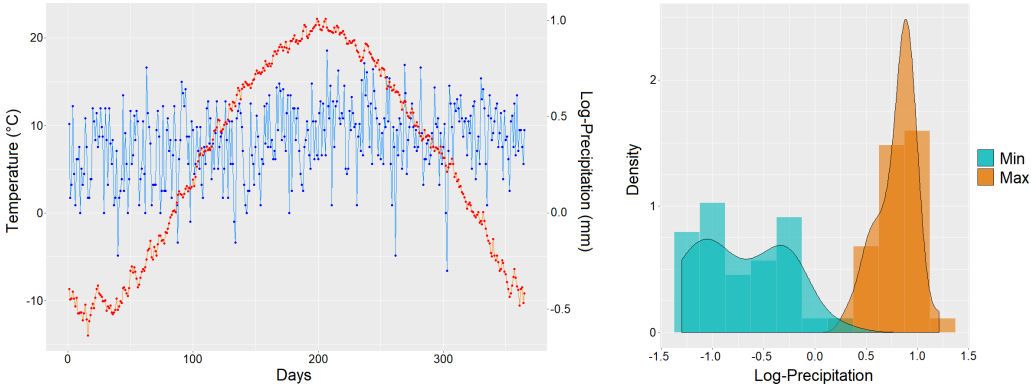


FIGURE 6. Temperature and log-precipitation measured over a year in Montreal (left). Empirical distribution of the minimal and maximal log-precipitation for the 35 weather stations (right).

Some authors (see *e.g.* [24]) have already highlighted the pertinence of using the matrix L defined in (3.2) in this dataset, because the predictors are ordered temporally so that the selection of isolated days instead of relevant sequences of days seems an unreliable procedure for statistical interpretation. To assess the models, we repeat $N = 100$ times the following experiment: $n_t = 25$ observations are randomly selected for calibration (*via* 2-fold

cross-validation) and estimation, the remaining $n_v = 10$ observations are used to compute the MSPE (3.1) related to the prediction of the minimum ($\min_p$) and maximum ($\max_p$) precipitation. We can see on Figure 7 that all structured PGGM perform almost identically, with the phenomenon described in the previous section still visible but to a lesser extent. We can even notice that structuring is hardly beneficial for this dataset, from a purely numerical point of view. This conclusion can also be found in [24], where the author compares the structured Elastic-Net with unstructured alternatives to predict the 0.25-, 0.50- and 0.75-quantiles of the log-precipitation, through independent regressions. But we will see that, in terms of variable selection and statistical interpretation, L and β still have a substancial role to play.

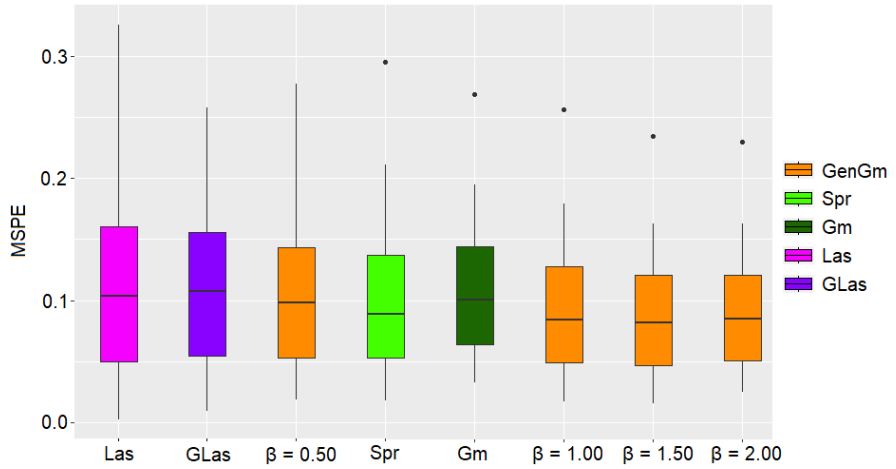


FIGURE 7. Mean squared prediction error for $N = 100$ repetitions of the experiment. GenGm for $\beta \in \{0.5, 1, 1.5, 2\}$ is compared with Spr, Gm, Las and GLas.

The point is that we have observed that the best prediction error does not usually coincide with a sparse solution (see Remark 3.3 above) when the coefficients or the covariates are not very contrasting. In particular, this was the case of our simulation study with $\pm\frac{1}{2}$ coefficients and $\mathcal{N}(0,1)$ covariates. So, just as they look at the Lasso's regularization paths, practitioners may choose the desired degree of sparsity, depending on p/n, by playing with the hyperparameters. Here, on the basis of the MSPE, most of the time we must retain $\mu \ll 10^{-2}$ and only a few direct links are set to zero. To look for sequences of days directly related to $\min_p$ and $\max_p$, we decided to constraint $\mu \geqslant 10^{-2}$ and focus on variable selection. The active set of Ω_{yx} is evaluated on the basis of $n_t = 25$ randomly chosen observations. The experiment is repeated $N = 100$ times, and the locations having a frequency of occurrence that exceeds 0.5 are retained (or, equivalently, those whose estimates have a non-zero median). This can be seen as a measure of variable importance. The results are given on Figures 8 and 9 below for $\min_p$ and $\max_p$, respectively, with a fixed set of regularization parameters and increasing values of β. The objective is to show the influence of the latter, all other things being equal. The colored areas highlight the days having a frequency of occurrence, represented by gray crosses, that exceeds 0.5 in the $N = 100$ repetitions of the experiment. Note that, since we retain $\lambda = 0$ in these experiments, GenGm for $\beta = 1$ coincides with Spr. We can see that the increasing pressure exerted by β on the estimation procedure tends to refine the selection by

giving priority to the most important variables and by dropping the others much more easily, at the cost of prediction results: we are undoubtedly in a selection process. The sequence of inclusions

$$\widehat{S}_{\beta_2} \subset \widehat{S}_{\beta_1} \quad \text{for } \beta_1 < \beta_2$$

that we observe for the estimated active sets is clearly a guarantee of quality for the selected variables.

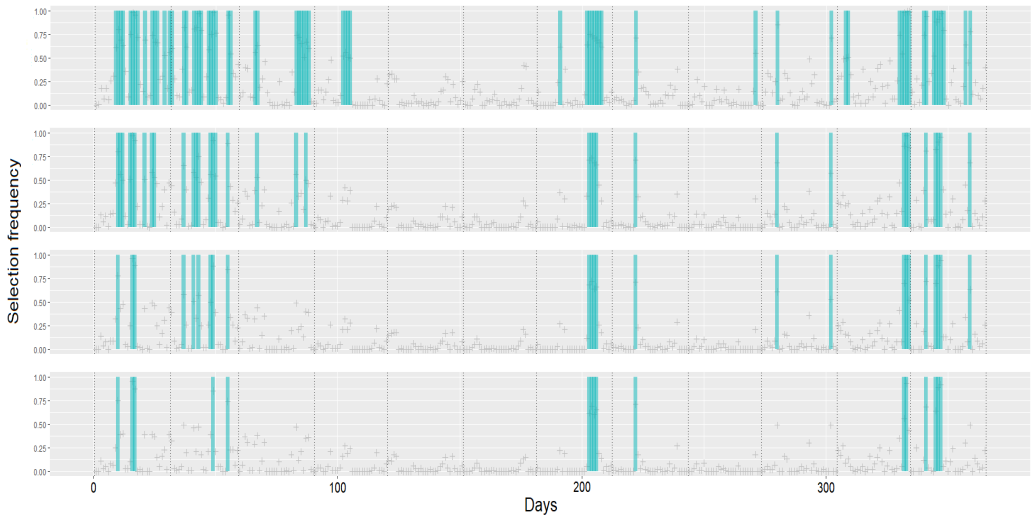


FIGURE 8. Variable selection for $\min_p$ by GenGm with $(\lambda, \mu, \eta) = (0, 0.05, 1)$ and, from top to bottom, $\beta \in \{0.5, 1, 1.5, 2\}$.

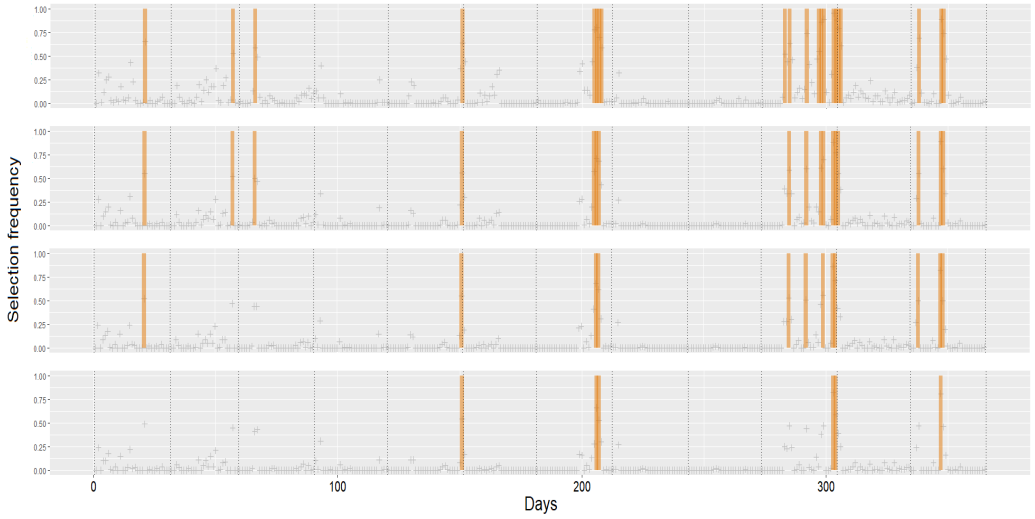


FIGURE 9. Variable selection for $\max_p$ by GenGm with $(\lambda, \mu, \eta) = (0, 0.05, 1)$ and, from top to bottom, $\beta \in \{0.5, 1, 1.5, 2\}$.

The median values of the estimated direct links between the temperature of the days and the pair $(\min_p, \max_p)$ are represented on Figure 10 together with the estimated regression

coefficients, for $\beta = 2$. We recall that the relation $B = -\Omega_{yx}^t \, \Omega_{yy}^{-1}$ simply lead to

$$\widehat{B} = -\widehat{\Omega}_{yx}^t \, \widehat{\Omega}_{yy}^{-1}.$$

We detect sequences of influent days in November, December, January and February, especially related to $\min_p$, positively at the end of the year and negatively at the beggining. This is broadly consistent with the analysis of [24] – even if the responses are not extremes but quantiles in it – with however two differences: the regression coefficients associated with $\max_p$ are much lower compared to $\min_p$ whereas it is not that clear in the reference, and an activity is also detected between July and August. The main explanation, at least for the first of them, probably lies in the use of graphical models that take into account the correlation between responses. Indeed, as can be seen on Figure 11 which gives an overview of the estimation of R obtained from the repeated experiments, a non-zero correlation is detected between the responses (≈ 0.32). The influence of November and December on all quantiles and that of January and February on the 0.75-quantile in [24] might actually be an artificial effect of the correlation with the 0.25-quantile. This is what our study suggests by highlighting $\min_p$ compared to $\max_p$: the 'real' effect appears to be on $\min_p$ whereas $\max_p$ seems to react only through a phenomenon of correlation with $\min_p$. From this point of view, the interest of graphical models instead of independent regressions is particularly obvious.

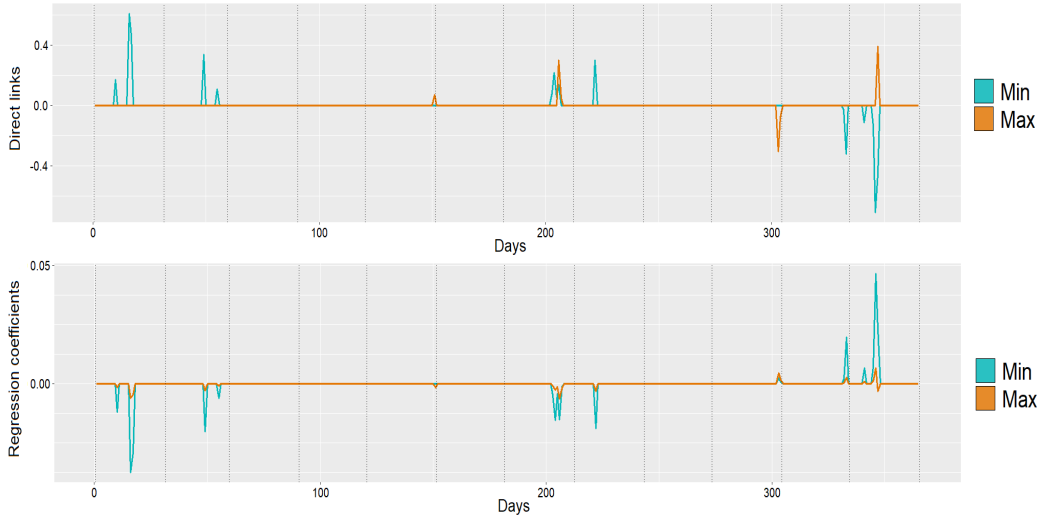


FIGURE 10. Estimated direct links (top) and regression coefficients (bottom) for the pair $(\min_p, \max_p)$ by GenGm with $(\lambda, \mu, \eta) = (0, 0.05, 1)$ and $\beta = 2$, after the $N = 100$ experiments. Dotted lines divide the panel into months.

Let us also mention that, interestingly enough, we notice that the role of η tends to depreciate for the large values of β. For example, for the same regularization parameters $(\lambda, \mu) = (0, 0.05)$ and $\beta = 2$, the difference between the estimated active sets for $\eta = 0.1$ and $\eta = 1$ is almost negligible (depending on the experiments, between 1 and 3 days are concerned, on average). Based on these studies and observations, we might conclude that β is insignificant when we are interested in the best prediction error on a validation set (even counterproductive with respect to computational times, *e.g.* compared to Spr), whereas it

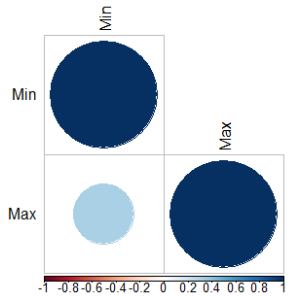


FIGURE 11. Estimated correlation between $\min_{\mathrm{p}}$ and $\max_{\mathrm{p}}$ by GenGm with $(\lambda, \mu, \eta) = (0, 0.05, 1)$ and $\beta = 2$, after the $N = 100$ experiments. The off-diagonal entry is approximately 0.32.

seems to have a substancial role to play when focusing on selection, by accelerating the discrimination of variables. In the first case, η has to be carefully adjusted while in the second case, β will quickly help to reach the desired sparsity.

Remark 3.4 (Structure matrix). For the simulations and the real dataset, we have used the popular first finite difference operator given in (3.2). Other examples can be found in the literature, like the promotion of a genetic distance for genomic selection in *Brassica napus* [7] or the bidimensional discretization of the Laplacian to work on handwritten digit recognition [24]. More generally, L can be used in a classic Bayesian prior supposed to promote some covariance structure on the direct links, with no 'physical' structuring in mind (like temporal, spatial or genetic proximity).

4. Conclusion

In conclusion, our work is a generalization of [29], using the same technical tools to establish an upper bound on the estimation error when a prior on the direct links generates an additional structural penalty in the objective, provided that the model is suitably regularized. Our work is also an improvement of [7] since, while being inspired by the methodology of the authors, we generalize the prior and give some theoretical guarantees. The empirical study shows that the hyperparametrization in the prior, although more expensive in adjusting the parameters, is likely to refine the selection results but clearly, this does not appear as a crucial improvement compared to the two previous points. Let us conclude the paper by highlighting two weaknesses that might be trails for future studies. On the one hand, the Laplace distribution is often used as a prior in the Bayesian Lasso (see *e.g.* Sec. 6.1 of [11]). However, our reasonings do not allow $\beta = 1/2$, which may correspond to a multivariate Laplace distribution on the direct links. Combined with the first finite difference operator L, the choice $\beta = 1/2$ could generate a Fused-Lasso-type penalty. In this regard, it would be challenging and interesting to obtain some theoretical guarantees for $\beta \geqslant 1/2$ and not only for $\beta \geqslant 1$, even if our probably too brief simulation study does not encourage the choice of $\beta < 1$. On the other hand, $\lambda = 0$ is a natural choice when q is small (this is in particular the configuration of [7]), not to mention that it is computionally faster. But, the proof of our theorem needs $\lambda > c_\lambda \, h_a > 0$ to hold. We think that a reasoning enabling to deal with $\lambda = 0$ should also be beneficial to the study. More generally, it would be instructive to consider a

very high-dimensional setting ($p \gg n$ and not only $p \sim 10^2$ although always larger than n, as in our experiments). Such studies should follow with omic data.

Acknowledgements and Fundings. The authors warmly thank the two anonymous reviewers for the careful reading and for making numerous useful corrections to improve the paper. We thank ALM (Angers Loire Métropole) and the ICO (Institut de Cancérologie de l'Ouest) for the financial support. This work is partially financed through the ALM grant and the "Programme opérationnel régional FEDER-FSE Pays de la Loire 2014-2020" noPL0015129 (EPICURE). The authors also thank Mario Campone (project leader and director of the ICO), Mathilde Colombié (scientific coordinator of EPICURE clinical trial) and Fadwa Ben Azzouz, biomathematician in Bioinfomics, for the initiation, the coordination and the smooth running of the project.

5. Technical proofs

We start in a first part by some useful linear algebra lemmas that will be repeatedly used subsequently, well-known for most of them. In a second part, we prove the joint convexity of the objective and our main result.

5.1. **Linear algebra.**

Lemma 5.1. *Let $A \in \mathbb{S}_+^d$ and $U \in \mathbb{R}^{d\times \ell}$. Then, $U^t A U \in \mathbb{S}_+^\ell$.*

Proof. Since A is symmetric with non-negative eigenvalues, there is an orthogonal matrix P such that $A = PDP^t$ with $D = \mathrm{diag}(\mathrm{sp}(A)) \in \mathbb{S}_+^d$. Thus, for all $v \in \mathbb{R}^\ell$, it follows that $\langle v, U^t A\, U\, v\rangle = \|D^{1/2}\, P^t\, U\, v\|^2 \geqslant 0$. □

Lemma 5.2. *Let $A \in \mathbb{S}_{++}^d$ and $B \in \mathbb{S}_+^d$. Then for all i, $\lambda_i(AB) \geqslant 0$.*

Proof. The equality $AB = A^{1/2}\,(A^{1/2}\, B\, A^{1/2})\, A^{-1/2}$ shows that AB and $A^{1/2}\, B\, A^{1/2}$ are similar, so they must share the same eigenvalues. From Lemma 5.1, $\lambda_i(A^{1/2}\, B\, A^{1/2}) \geqslant 0$. □

Lemma 5.3. *Let $A \in \mathbb{S}_+^d$ and $B \in \mathbb{S}_+^d$. Then,*

$$\lambda_{\min}(A)\,\mathrm{tr}(B) \;\leqslant\; \mathrm{tr}(AB) \;\leqslant\; \lambda_{\max}(A)\,\mathrm{tr}(B).$$

Proof. Since $A - \lambda_{\min}(A) I_d \in \mathbb{S}_+^d$ and $B \in \mathbb{S}_+^d$,

$$\mathrm{tr}((A - \lambda_{\min}(A) I_d)\, B) = \mathrm{tr}(B^{1/2}\,(A - \lambda_{\min}(A) I_d)\, B^{1/2}) \geqslant 0$$

from Lemma 5.1, thus $\mathrm{tr}(AB) \geqslant \lambda_{\min}(A)\,\mathrm{tr}(B)$. The other inequality is obtained through $\lambda_{\max}(A) I_d - A \in \mathbb{S}_+^d$. □

Lemma 5.4. *Let $A \in \mathbb{S}_{++}^d$ and $B \in \mathbb{S}_+^d$. Then,*

$$\lambda_{\min}(A)\,\lambda_{\min}(B) \;\leqslant\; \lambda_{\min}(AB) \quad \textit{and} \quad \lambda_{\max}(AB) \;\leqslant\; \lambda_{\max}(A)\,\lambda_{\max}(B).$$

Proof. On the one hand, $\lambda_{\max}(AB) \leqslant \|AB\|_2 \leqslant \|A\|_2\, \|B\|_2 = \lambda_{\max}(A)\,\lambda_{\max}(B)$, since A and B are symmetric and since, from Lemma 5.2 and by hypothesis, all eigenvalues appearing in the relation are non-negative. Suppose now that B is invertible so that both A^{-1} and B^{-1} belong to $\mathbb{S}_{++}^d$. Then, $\lambda_{\max}((AB)^{-1}) \leqslant \lambda_{\max}(A^{-1})\,\lambda_{\max}(B^{-1})$ and this immediately gives $\lambda_{\min}(AB) \geqslant \lambda_{\min}(A)\,\lambda_{\min}(B)$. If B is not invertible, the relation trivially holds since we still have $\lambda_{\min}(AB) \geqslant 0$ from Lemma 5.2. □

Lemma 5.5. *Let $A \in \mathbb{S}_+^d$ and $U \in \mathbb{R}^{d\times\ell}$. Then,*

$$\lambda_{\min}(A)\,\|U\|_F^2 \leqslant \operatorname{tr}(U^t A U) \leqslant \lambda_{\max}(A)\,\|U\|_F^2.$$

Proof. Denote by u_i the i-th column of U. It is not hard to see that the i-th diagonal element of $U^t A U$ is $u_i^t A\,u_i \geqslant \lambda_{\min}(A)\,\|u_i\|^2 \geqslant 0$. Thus,

$$\operatorname{tr}(U^t A U) = \sum_{i=1}^{\ell} u_i^t A\,u_i \geqslant \lambda_{\min}(A) \sum_{i=1}^{\ell} \|u_i\|^2 = \lambda_{\min}(A)\,\|U\|_F^2.$$

The upper bound stems from $0 \leqslant u_i^t A\,u_i \leqslant \lambda_{\max}(A)\,\|u_i\|^2$. □

Lemma 5.6. *Let A and B be symmetric matrices of same dimensions. Then,*

$$\lambda_{\min}(A) + \lambda_{\min}(B) \leqslant \lambda_{\min}(A+B) \quad \text{and} \quad \lambda_{\max}(A+B) \leqslant \lambda_{\max}(A) + \lambda_{\max}(B).$$

Proof. These are just two special cases of Weyl inequalities. We refer the reader to Thm. 4.3.1 of [12], for example. □

5.2. **Convexity of the objective.** We know from Prop. 1 of [29] and the convexity of the elementwise ℓ_1 norm that $L_n(\Omega_{yy},\Omega_{yx}) - \eta\,\langle\!\langle L, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx}\rangle\!\rangle^\beta$ is itself convex, but it remains to show that this is still the case with the additional smooth penalty.

Proof of Proposition 2.1. Recall that $\Theta = \mathbb{S}_{++}^q \times \mathbb{R}^{q\times p}$ and consider the mapping $\Phi : \Theta \to \mathbb{S}_+^p$ defined as

$$\forall\,(A,B) \in \Theta, \quad \Phi(A,B) = B^t A^{-1} B.$$

We can already note from Lemma 5.1 that $\operatorname{tr}(\Phi(A,B)) \geqslant 0$. Moreover, for all $0 \leqslant h \leqslant 1$ and all $Z_i = (A_i, B_i) \in \Theta$, $i = 1, 2$, it is easy to see that

$$S_h(Z_1,Z_2) = h\,\Phi(Z_1) + (1-h)\,\Phi(Z_2) - \Phi(hZ_1 + (1-h)Z_2) \tag{5.1}$$

is the Schur complement of $hA_1 + (1-h)A_2$ in the matrix

$$M_h(Z_1,Z_2) = h\begin{pmatrix} A_1 & B_1 \\ B_1^t & B_1^t\,A_1^{-1}B_1 \end{pmatrix} + (1-h)\begin{pmatrix} A_2 & B_2 \\ B_2^t & B_2^t\,A_2^{-1}B_2 \end{pmatrix}. \tag{5.2}$$

But the decomposition

$$\begin{pmatrix} A^{1/2} & A^{-1/2}\,B \\ 0 & 0 \end{pmatrix}^t \begin{pmatrix} A^{1/2} & A^{-1/2}\,B \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} A & B \\ B^t & B^t\,A^{-1}B \end{pmatrix}$$

directly shows that $M_h(Z_1,Z_2)$ in (5.2) is symmetric and positive semi-definite. It is well-known (see *e.g.* Appendix A.5.5 of [4]) that in that case, the Schur complement (5.1) must also be positive semi-definite. Consequently, for $Z_i = (\Omega_{i,yy}, \Omega_{i,yx}L^{1/2})$, $i = 1, 2$, taking the trace of $S_h(Z_1,Z_2)$ and considering $\beta \geqslant 1$,

$$\begin{aligned}
\langle\!\langle L, P_h^t\,Q_h^{-1}P_h\rangle\!\rangle^\beta &= (\operatorname{tr}(\Phi(hZ_1 + (1-h)Z_2)))^\beta \\
&\leqslant (h\operatorname{tr}(\Phi(Z_1)) + (1-h)\operatorname{tr}(\Phi(Z_2)))^\beta \\
&= (h\,\langle\!\langle L, \Omega_{1,yx}^t\,\Omega_{1,yy}^{-1}\,\Omega_{1,yx}\rangle\!\rangle + (1-h)\,\langle\!\langle L, \Omega_{2,yx}^t\,\Omega_{2,yy}^{-1}\,\Omega_{2,yx}\rangle\!\rangle)^\beta \\
&\leqslant h\,\langle\!\langle L, \Omega_{1,yx}^t\,\Omega_{1,yy}^{-1}\,\Omega_{1,yx}\rangle\!\rangle^\beta + (1-h)\,\langle\!\langle L, \Omega_{2,yx}^t\,\Omega_{2,yy}^{-1}\,\Omega_{2,yx}\rangle\!\rangle^\beta
\end{aligned}$$

where $P_h = h\,\Omega_{1,yx} + (1-h)\,\Omega_{2,yx}$ and $Q_h = h\,\Omega_{1,yy} + (1-h)\,\Omega_{2,yy}$. This convexity inequality concludes the proof. □

5.3. **Theoretical guarantees.**

Proof of Theorem 2.1. Let $R_n(\theta)$ be the the smooth part of the objective (2.2),

$$\begin{aligned} R_n(\theta) &= -\ln\det(\Omega_{yy}) + \langle\!\langle S_{yy}^{(n)}, \Omega_{yy}\rangle\!\rangle + 2\,\langle\!\langle S_{yx}^{(n)}, \Omega_{yx}\rangle\!\rangle \\ &\qquad + \langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx}\rangle\!\rangle + \eta\,\langle\!\langle L, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx}\rangle\!\rangle^{\beta}. \end{aligned} \tag{5.3}$$

For any $\theta \in \Theta$ and $t \in \mathbb{R}$, by a Taylor expansion,

$$R_n(\theta^* + t\,(\theta - \theta^*)) = R_n(\theta^*) + t\,\langle\!\langle \nabla R_n(\theta^*), \theta - \theta^*\rangle\!\rangle + e_t(\theta, \theta^*) \tag{5.4}$$

for some second-order error term $e_t(\theta, \theta^*)$. Consider the reparametrization

$$\phi(t) = R_n(\theta^* + t\,(\theta - \theta^*)) \tag{5.5}$$

so that $\phi'(0) = \langle\!\langle \nabla R_n(\theta^*), \theta - \theta^*\rangle\!\rangle$. Let $\delta\theta_{yy} = \Omega_{yy} - \Omega_{yy}^*$ and $\delta\theta_{yx} = \Omega_{yx} - \Omega_{yx}^*$, let also $\delta\theta = \theta - \theta^*$ in a compact form. The estimation error is denoted

$$\delta\vartheta = \widehat{\theta} - \theta^* = (\widehat{\Omega}_{yy} - \Omega_{yy}^*, \widehat{\Omega}_{yx} - \Omega_{yx}^*) = (\delta\vartheta_{yy}, \delta\vartheta_{yx}). \tag{5.6}$$

Before we start the actual proof, some additional lemmas are needed. They constitute a local study in a sort of r^*-neighborhood of θ^* that we define as

$$N_{r,\alpha}(\theta^*) = \big\{\theta \in \Theta,\ \|\delta\theta\|_F \leqslant r^* \text{ and } |[\delta\theta]_{\bar{S}}|_1 \leqslant \alpha |[\delta\theta]_S|_1\big\}. \tag{5.7}$$

Our strategy can be summarized as follows:

- → (Lemma 5.9) Show that there exists a configuration for the regularization parameters (λ, μ, η) so that the estimation error satisfies $|[\delta\vartheta]_{\bar{S}}|_1 \leqslant \alpha |[\delta\vartheta]_S|_1$ for some $\alpha > 0$.
- → (Lemma 5.10) Find some $r^* > 0$ and $\gamma_{r,\eta,\beta,p} > 0$ such that $e_1(\theta, \theta^*) > \gamma_{r,\eta,\beta,p}\|\delta\theta\|_F^2$ as soon as $\theta \in N_{r,\alpha}(\theta^*)$.
- → (Lemma 5.11) Exploit this result to show that the estimation error must also satisfy $\|\delta\vartheta\|_F \leqslant r^*$ provided that $\max\{h_a, h_b\}$ is small enough.
- → (Lemma 5.12) Conclude that the theorem holds with high probability, provided that n is large enough.

For the sake of readability, we refer the reader to the Appendix for the numerous constants that are about to appear in the following lemmas and proofs. Thereafter, $N_{r,\alpha}(\theta^*)$ will always refer to α in (A.4) and r^* in (A.6), while the second hypothesis (H_2) given below is to be assumed with the smallest integer greater than s_α in (A.5). This is a random hypothesis, which will be controlled with a probability, related to the proximity between the empirical covariance and the true covariance of the predictors, since we recall that $S^{(n)}$ has no reason to be an excellent approximation of Σ^* when $p \gg n$. This is also assumed by the authors of [29], it is a kind of restricted isometry propertie (RIP), well-known in high-dimensional studies. In particular, we will see through Lemma 5.12 that it is satisfied with high probability provided that n is large enough.

$$(H_2) \qquad \forall\, u \neq 0 \text{ such that } |u|_0 \leqslant \lceil s_\alpha \rceil, \quad \frac{1}{2}\,u^t\,\Sigma_{xx}^*\,u \leqslant u^t\,S_{xx}^{(n)}\,u \leqslant \frac{3}{2}\,u^t\,\Sigma_{xx}^*\,u.$$

$$\text{In addition,} \quad \lambda_{\max}(\Omega_{yx}^*\,S_{xx}^{(n)}\,\Omega_{yx}^{*\,t}) \leqslant \frac{7}{5}\,\lambda_{\max}(\Omega_{yx}^*\,\Sigma_{xx}^*\,\Omega_{yx}^{*\,t}).$$

The next two lemmas give some bounds for expressions that will appear repeatedly.

Lemma 5.7. *Under* (H$_1$) *and* (H$_2$), *for all* $\theta \in N_{r,\alpha}(\theta^*)$, *we have the bound*

$$\lambda_{\max}(\Omega_{yy}^{-1}\,\Omega_{yx}\,S_{xx}^{(n)}\,\Omega_{yx}^t) \leqslant \overline{\omega}_S$$

where $\overline{\omega}_S$ *is given in* (A.1). *In addition,*

$$\operatorname{tr}(\delta\theta_{yx}\,S_{xx}^{(n)}\,\delta\theta_{yx}^t) \geqslant \frac{\lambda_{\min}(\Sigma_{xx}^*)}{10}\,\|\delta\theta_{yx}\|_F^2.$$

Proof. Similar reasonings may be found in the proofs of Lem. 1-2 of [29]. We simply reworked the constants to make them stick to our study. □

Lemma 5.8. *Under* (H$_1$), *for all* $\theta \in N_{r,\alpha}(\theta^*)$, *we have the bounds*

$$\lambda_{\min}(\Omega_{yy}^{-1}\,\Omega_{yx}\,L\,\Omega_{yx}^t) \geqslant \underline{\omega}_L \quad \textit{and} \quad \lambda_{\max}(\Omega_{yy}^{-1}\,\Omega_{yx}\,L\,\Omega_{yx}^t) \leqslant \overline{\omega}_L$$

where $\underline{\omega}_L$ *and* $\overline{\omega}_L$ *are given in* (A.1). *As a corollary,*

$$p\,\underline{\omega}_L \leqslant \langle\!\langle L, \Omega_{yx}^t\,\Omega_{yy}^{-1}\,\Omega_{yx}\rangle\!\rangle \leqslant p\,\overline{\omega}_L.$$

Proof. From Lemmas 5.1 and 5.6,

$$\begin{aligned}
2\,\lambda_{\min}(\Omega_{yx}\,L\,\Omega_{yx}^t) &\geqslant 2\left(\lambda_{\min}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t}) + \lambda_{\min}(\delta\theta_{yx}\,L\,\Omega_{yx}^{*t} + \Omega_{yx}^*\,L\,\delta\theta_{yx}^t)\right)\\
&\geqslant 2\left(\lambda_{\min}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t}) - \|\delta\theta_{yx}\,L\,\Omega_{yx}^{*t} + \Omega_{yx}^*\,L\,\delta\theta_{yx}^t\|_2\right)\\
&\geqslant 2\left(\lambda_{\min}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t}) - 2\,\|\delta\theta_{yx}\|_F\,\|L\,\Omega_{yx}^{*t}\|_2\right) \geqslant \lambda_{\min}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t})
\end{aligned}$$

as soon as $\|\delta\theta_{yx}\|_F \leqslant r^*$. From Lemma 5.4, we get

$$\lambda_{\min}(\Omega_{yy}^{-1}\,\Omega_{yx}\,L\,\Omega_{yx}^t) \geqslant \frac{\lambda_{\min}(\Omega_{yx}\,L\,\Omega_{yx}^t)}{\lambda_{\max}(\Omega_{yy})} \geqslant \frac{\lambda_{\min}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t})}{4\,\lambda_{\max}(\Omega_{yy}^*)}$$

where the inequality in the denominator comes from $\lambda_{\max}(\Omega_{yy}) \leqslant \lambda_{\max}(\Omega_{yy}^*) + \lambda_{\max}(\delta\theta_{yy})$, *via* Lemma 5.6, and the fact that $\lambda_{\max}(\delta\theta_{yy}) \leqslant \|\delta\theta_{yy}\|_F \leqslant r^* \leqslant \lambda_{\max}(\Omega_{yy}^*)$. For the upper bound, a similar logic gives, with Lemma 5.5,

$$\begin{aligned}
\sqrt{\lambda_{\max}(\Omega_{yx}\,L\,\Omega_{yx}^t)} &\leqslant \sqrt{\lambda_{\max}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t})} + \sqrt{\operatorname{tr}(\delta\theta_{yx}\,L\,\delta\theta_{yx}^t)}\\
&\leqslant \sqrt{\lambda_{\max}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t})} + \|\delta\theta_{yx}\|_F\,\sqrt{\lambda_{\max}(L)} \leqslant \sqrt{2\,\lambda_{\max}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t})}
\end{aligned}$$

for $\|\delta\theta_{yx}\|_F \leqslant r^*$. It follows from Lemma 5.4 that

$$\lambda_{\max}(\Omega_{yy}^{-1}\,\Omega_{yx}\,L\,\Omega_{yx}^t) \leqslant \frac{\lambda_{\max}(\Omega_{yx}\,L\,\Omega_{yx}^t)}{\lambda_{\min}(\Omega_{yy})} \leqslant \frac{4\,\lambda_{\max}(\Omega_{yx}^*\,L\,\Omega_{yx}^{*t})}{\lambda_{\min}(\Omega_{yy}^*)}$$

where the inequality in the denominator comes from $\lambda_{\min}(\Omega_{yy}) \geqslant \lambda_{\min}(\Omega_{yy}^*) + \lambda_{\min}(\delta\theta_{yy})$, *via* Lemma 5.6, and the fact that $2\,\lambda_{\min}(\delta\theta_{yy}) \geqslant -2\,\|\delta\theta_{yy}\|_F \geqslant -2\,r^* \geqslant -\lambda_{\min}(\Omega_{yy}^*)$. The corollary that concludes the lemma is now immediate. □

Lemma 5.9. *Assume that* λ, μ *and* η *are chosen according to the configuration of the theorem. Then, under* (H$_1$), *the estimation error satisfies*

$$|[\delta\vartheta]_{\bar{S}}|_1 \leqslant \alpha\,|[\delta\vartheta]_S|_1$$

where $\alpha > 0$ *is given in* (A.4).

Proof. Taking $t = 1$ in the Taylor expansion (5.4) with $\theta = \widehat{\theta}$ and considering the definition of ϕ in (5.5), by convexity,

$$R_n(\widehat{\theta}) - R_n(\theta^*) \geqslant \phi'(0).$$

The first derivative of ϕ will be explicitely computed in (5.11). For $t = 0$, we find

$$\begin{aligned}
\phi'(0) &= -\langle\!\langle \Omega_{yy}^{*\,-1}, \delta\vartheta_{yy} \rangle\!\rangle + \langle\!\langle S_{yy}^{(n)}, \delta\vartheta_{yy} \rangle\!\rangle + 2\,\langle\!\langle S_{yx}^{(n)}, \delta\vartheta_{yx} \rangle\!\rangle \\
&\quad + 2\,\langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^{*\,t}\,\Omega_{yy}^{*\,-1}\,\delta\vartheta_{yx} \rangle\!\rangle - \langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^{*\,t}\,\Omega_{yy}^{*\,-1}\,\delta\vartheta_{yy}\,\Omega_{yy}^{*\,-1}\,\Omega_{yx}^{*} \rangle\!\rangle \\
&\quad + \eta\beta\, s_L^{\beta-1}\big[2\,\langle\!\langle L, \Omega_{yx}^{*\,t}\,\Omega_{yy}^{*\,-1}\,\delta\vartheta_{yx} \rangle\!\rangle - \langle\!\langle L, \Omega_{yx}^{*\,t}\,\Omega_{yy}^{*\,-1}\,\delta\vartheta_{yy}\,\Omega_{yy}^{*\,-1}\,\Omega_{yx}^{*} \rangle\!\rangle\big] \\
&= \langle\!\langle A_n + \eta\beta\, s_L^{\beta-1} C_A, \delta\vartheta_{yy} \rangle\!\rangle + \langle\!\langle B_n + \eta\beta\, s_L^{\beta-1} C_B, \delta\vartheta_{yx} \rangle\!\rangle
\end{aligned}$$

where s_L is given in (A.3), where, through the blockwise relations (1.3), we recognize the random matrices A_n (with max norm h_a) and B_n (with max norm h_b) defined in (2.3) and (2.4), and where, coming from the structural regularization term,

$$C_A = -\Omega_{yy}^{*\,-1}\,\Omega_{yx}^{*}\,L\,\Omega_{yx}^{*\,t}\,\Omega_{yy}^{*\,-1} \quad \text{and} \quad C_B = 2\,\Omega_{yy}^{*\,-1}\,\Omega_{yx}^{*}\,L.$$

Whence it follows from the well-known relation $|\mathrm{tr}(M_1 M_2)| \leqslant |M_1|_\infty\,|M_2|_1$, where M_1 and M_2 are compatible matrices, that

$$\phi'(0) \geqslant -\frac{\lambda}{c_\lambda}\,|\delta\vartheta_{yy}|_1 - \eta\beta\, s_L^{\beta-1}\ell_a\,|\delta\vartheta_{yy}|_1 - \frac{\mu}{c_\mu}\,|\delta\vartheta_{yx}|_1 - \eta\beta\, s_L^{\beta-1}\ell_b\,|\delta\vartheta_{yx}|_1$$

making use of the constants (A.3), $\lambda \geqslant c_\lambda\, h_a$ and $\mu \geqslant c_\mu\, h_b$. For the sake of clarity, let

$$\Delta_n(\theta, \theta^*) = R_n(\theta) + \lambda\,|\Omega_{yy}|_1^- + \mu\,|\Omega_{yx}|_1 - R_n(\theta^*) - \lambda\,|\Omega_{yy}^*|_1^- - \mu\,|\Omega_{yx}^*|_1.$$

For all $\theta \in \Theta$,

$$\begin{aligned}
|\Omega_{yy}|_1^- - |\Omega_{yy}^*|_1^- &= |[\Omega_{yy}^* + \delta\theta_{yy}]_S|_1^- + |[\delta\theta_{yy}]_{\bar{S}}|_1^- - |[\Omega_{yy}^*]_S|_1^- \\
&\geqslant \big||[\Omega_{yy}^*]_S|_1^- - |[\delta\theta_{yy}]_S|_1^-\big| + |[\delta\theta_{yy}]_{\bar{S}}|_1^- - |[\Omega_{yy}^*]_S|_1^- \\
&\geqslant |[\delta\theta_{yy}]_{\bar{S}}|_1 - |[\delta\theta_{yy}]_S|_1
\end{aligned}$$

from the triangle inequality and the fact that, as Ω_{yy}^* is positive definite, the diagonal must belong to S, *i.e.* $(j, j) \in S$ for all $1 \leqslant j \leqslant q$ so that any square matrix M of size q is such that $[M]_{\bar{S}}$ has diagonal elements all equal to zero. A similar bound obviously holds for $|\Omega_{yx}|_1 - |\Omega_{yx}^*|_1$. Now, a straightforward calculation shows that

$$\Delta_n(\widehat{\theta}, \theta^*) \geqslant \underline{c}\,\big(|[\delta\vartheta_{yy}]_{\bar{S}}|_1 + |[\delta\vartheta_{yx}]_{\bar{S}}|_1\big) - \overline{c}\,\big(|[\delta\vartheta_{yy}]_S|_1 + |[\delta\vartheta_{yx}]_S|_1\big) \tag{5.8}$$

where

$$\overline{c} = \max\left\{\frac{(c_\lambda + 1)\lambda}{c_\lambda} + \eta\beta\, s_L^{\beta-1}\ell_a, \frac{(c_\mu + 1)\mu}{c_\mu} + \eta\beta\, s_L^{\beta-1}\ell_b\right\}$$

and

$$\underline{c} = \min\left\{\frac{(c_\lambda - 1)\lambda}{c_\lambda} - \eta\beta\, s_L^{\beta-1}\ell_a, \frac{(c_\mu - 1)\mu}{c_\mu} - \eta\beta\, s_L^{\beta-1}\ell_b\right\}.$$

Thus, provided that $\underline{c} > 0$, which is stated in the configuration of the theorem, it only remains to note that, necessarily,

$$\Delta_n(\widehat{\theta}, \theta^*) \leqslant 0$$

since $\widehat{\theta}$ is the global minimizer of $\theta \mapsto R_n(\theta) + \lambda\,|\Omega_{yy}|_1^- + \mu\,|\Omega_{yx}|_1$. The identification of α given in (A.4) easily follows. □

Lemma 5.10. *Under* (H_1) *and* (H_2)*, the second-order error term of* (5.4) *satisfies, for* $t = 1$ *and all* $\theta \in N_{r,\alpha}(\theta^*)$*,*

$$e_1(\theta, \theta^*) > \gamma_{r,\eta,\beta,p} \, \|\delta\theta\|_F^2$$

where $\gamma_{r,\eta,\beta,p} > 0$ *is given in* (A.7)*.*

Proof. From the definition of ϕ in (5.5) and the fact that $\phi'(0) = \langle\!\langle \nabla R_n(\theta^*), \theta - \theta^* \rangle\!\rangle$, there exists $h \in \,]0, 1[$ satisfying

$$e_1(\theta, \theta^*) = \frac{1}{2}\,\phi''(h). \tag{5.9}$$

To simplify the calculations, let

$$u_L = \langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle. \tag{5.10}$$

We are going to study the behavior of $R_n(\Omega_{yy}, \Omega_{yx})$ in the directions $\Omega_{yy} = \Omega_{yy}^* + t\,\delta\theta_{yy}$ and $\Omega_{yx} = \Omega_{yx}^* + t\,\delta\theta_{yx}$ through $\phi(t)$, where we recall that $\delta\theta_{yy} = \Omega_{yy} - \Omega_{yy}^*$ and $\delta\theta_{yx} = \Omega_{yx} - \Omega_{yx}^*$. One can see that $\phi(t)$ moves from $R_n(\Omega_{yy}, \Omega_{yx})$ to $R_n(\Omega_{yy}^*, \Omega_{yx}^*)$ as t decreases from 1 to 0. The first derivative is

$$\begin{aligned}
\phi'(t) \;=\; & -\langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \rangle\!\rangle + \langle\!\langle S_{yy}^{(n)}, \delta\theta_{yy} \rangle\!\rangle + 2\,\langle\!\langle S_{yx}^{(n)}, \delta\theta_{yx} \rangle\!\rangle \\
& + 2\,\langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle - \langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \\
& + \eta\beta \, u_L^{\beta-1} \big[2\,\langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle - \langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \big].
\end{aligned} \tag{5.11}$$

The second derivative is tedious to write but straightforward to establish,

$$\begin{aligned}
\phi''(t) \;=\; & \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle + 2\,\big[\langle\!\langle S_{xx}^{(n)}, \delta\theta_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle - 2\,\langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle \\
& + \langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \big] \\
& + 2\,\eta\beta \, u_L^{\beta-1} \big[\langle\!\langle L, \delta\theta_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle - 2\,\langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle \\
& + \langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \big] \\
& + \eta\beta(\beta - 1) \, u_L^{\beta-2} \big[2\,\langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yx} \rangle\!\rangle - \langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \big]^2.
\end{aligned} \tag{5.12}$$

First, from the combination of Lemmas 5.1 and 5.8, we clearly have $u_L \geqslant 0$. We also note that $0 \leqslant \|\frac{2}{c} M_1 - c M_2\|_F^2 = \frac{4}{c^2}\,\|M_1\|_F^2 - 4\,\langle\!\langle M_1, M_2 \rangle\!\rangle + c^2\,\|M_2\|_F^2$ for any $c \neq 0$ and any matrices M_1 and M_2 of same dimensions. It follows, after some reorganizations, that for any $c \neq 0$ and $d \neq 0$,

$$\begin{aligned}
\phi''(t) \;\geqslant\; & \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle \\
& + c_1\,\langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, S_{xx}^{(n)} \, \delta\theta_{yx}^t \rangle\!\rangle + c_2\,\langle\!\langle S_{xx}^{(n)}, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \\
& + \eta\beta \, u_L^{\beta-1} \big[d_1\,\langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, L \, \delta\theta_{yx}^t \rangle\!\rangle + d_2\,\langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \big]
\end{aligned}$$

where $c_1 = 2 - \frac{4}{c^2}$, $c_2 = 2 - c^2$, $d_1 = 2 - \frac{4}{d^2}$ and $d_2 = 2 - d^2$. Here we exploited the previous inequality twice, $u_L \geqslant 0$ and $\beta \geqslant 1$. From Lemmas 5.1, 5.3, 5.7 and 5.8, using $\mathrm{sp}(M_1\,M_2) = \mathrm{sp}(M_2\,M_1)$ for square matrices M_1 and M_2, we obtain

$$\langle\!\langle L, \Omega_{yx}^t \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \Omega_{yx} \rangle\!\rangle \leqslant \overline{\omega}_L \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle$$

where $\overline{\omega}_L$ is defined in (A.1). Replacing L by $S_{xx}^{(n)}$ and $\overline{\omega}_L$ by $\overline{\omega}_S$, a similar bound obviously holds. Suppose that c and d are chosen so that $c_1 > 0$, $d_1 > 0$, $c_2 < 0$ and $d_2 < 0$. Then,

$$\phi''(t) \;\geqslant\; \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle \, \big[1 - |c_2|\,\overline{\omega}_S - \eta\beta \, u_L^{\beta-1} |d_2|\,\overline{\omega}_L \big]$$

$$
\begin{aligned}
&\quad + c_1 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, S_{xx}^{(n)} \, \delta\theta_{yx}^t \rangle\!\rangle + \eta\beta \, u_L^{\beta-1} d_1 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, L \, \delta\theta_{yx}^t \rangle\!\rangle \\
\geqslant \;& \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle \left[1 - |c_2| \, \overline{\omega}_S - \eta\beta \, (p \, \overline{\omega}_L)^{\beta-1} |d_2| \, \overline{\omega}_L \right] \\
&\quad + c_1 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, S_{xx}^{(n)} \, \delta\theta_{yx}^t \rangle\!\rangle + \eta\beta \, (p \, \underline{\omega}_L)^{\beta-1} d_1 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, L \, \delta\theta_{yx}^t \rangle\!\rangle.
\end{aligned}
$$

Now choose $\epsilon_S > 0$ and $\epsilon_L > 0$ small enough so that $\epsilon_S \, \overline{\omega}_S + \eta\beta \, p^{\beta-1} \, \overline{\omega}_L^{\beta} \, \epsilon_L < 1$ and fix $c = \sqrt{2 + \epsilon_S}$ and $d = \sqrt{2 + \epsilon_L}$. We finally obtain

$$
\phi''(t) \geqslant a_1 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle + a_2 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, S_{xx}^{(n)} \, \delta\theta_{yx}^t \rangle\!\rangle + a_3 \, \langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, L \, \delta\theta_{yx}^t \rangle\!\rangle \tag{5.13}
$$

where these positive constants are respectively given by

$$
a_1 = 1 - \epsilon_S \, \overline{\omega}_S - \eta\beta \, p^{\beta-1} \, \overline{\omega}_L^{\beta} \, \epsilon_L, \quad a_2 = \frac{2 \, \epsilon_S}{2 + \epsilon_S} \quad \text{and} \quad a_3 = \eta\beta \, (p \, \underline{\omega}_L)^{\beta-1} \frac{2 \, \epsilon_L}{2 + \epsilon_L}.
$$

The combination of Lemmas 5.1, 5.3 and 5.5 gives, uniformly in $t \in [0, 1]$,

$$
\langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy} \rangle\!\rangle \geqslant \lambda_{\min}(\Omega_{yy}^{-1}) \, \mathrm{tr}(\delta\theta_{yy} \, \Omega_{yy}^{-1} \, \delta\theta_{yy}) \geqslant \frac{\|\delta\theta_{yy}\|_F^2}{4 \, \lambda_{\max}^2(\Omega_{yy}^*)}
$$

where the inequality in the denominator comes from $\lambda_{\max}(\Omega_{yy}) \leqslant 2 \, \lambda_{\max}(\Omega_{yy}^*)$ already established in the proof of Lemma 5.8. Similarly,

$$
\langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yx} \, L \, \delta\theta_{yx}^t \rangle\!\rangle \geqslant \lambda_{\min}(\Omega_{yy}^{-1}) \, \mathrm{tr}(\delta\theta_{yx} \, L \, \delta\theta_{yx}^t) \geqslant \frac{\lambda_{\min}(L) \, \|\delta\theta_{yx}\|_F^2}{2 \, \lambda_{\max}(\Omega_{yy}^*)}.
$$

Lemma 5.7 directly enables to bound the last term,

$$
\langle\!\langle \Omega_{yy}^{-1}, \delta\theta_{yy} \, S_{xx}^{(n)} \, \delta\theta_{yy} \rangle\!\rangle \geqslant \lambda_{\min}(\Omega_{yy}^{-1}) \, \mathrm{tr}(\delta\theta_{yx} \, S_{xx}^{(n)} \, \delta\theta_{yx}^t) \geqslant \frac{\lambda_{\min}(\Sigma_{xx}^*) \, \|\delta\theta_{yx}\|_F^2}{20 \, \lambda_{\max}(\Omega_{yy}^*)}.
$$

In conclusion, combining (5.9), (5.13) and the upper bounds above,

$$
\begin{aligned}
e_1(\theta, \theta^*) \;&\geqslant\; \frac{a_1 \, \|\delta\theta_{yy}\|_F^2}{8 \, \lambda_{\max}^2(\Omega_{yy}^*)} + \frac{a_2 \, \lambda_{\min}(L) \, \|\delta\theta_{yx}\|_F^2}{4 \, \lambda_{\max}(\Omega_{yy}^*)} + \frac{a_3 \, \lambda_{\min}(\Sigma_{xx}^*) \, \|\delta\theta_{yx}\|_F^2}{40 \, \lambda_{\max}(\Omega_{yy}^*)} \\
&\geqslant\; \min \left\{ \frac{a_1}{8 \, \lambda_{\max}^2(\Omega_{yy}^*)}, \frac{a_2 \, \lambda_{\min}(L)}{4 \, \lambda_{\max}(\Omega_{yy}^*)} + \frac{a_3 \, \lambda_{\min}(\Sigma_{xx}^*)}{40 \, \lambda_{\max}(\Omega_{yy}^*)} \right\} \|\delta\theta\|_F^2
\end{aligned}
$$

and we clearly identify $\gamma_{r,\eta,\beta,p} > 0$. □

Lemma 5.11. *Assume that λ, μ and η are chosen according to the configuration of the theorem. Suppose also that h_a in* (2.3) *and h_b in* (2.4) *satisfy*

$$
\max\{h_a, h_b\} < \frac{r^* \, \gamma_{r,\eta,\beta,p}}{c_{\lambda,\mu} \sqrt{|S|}}
$$

where r^* is given in (A.6)*, $\gamma_{r,\eta,\beta,p}$ in* (A.7) *and $c_{\lambda,\mu}$ in* (A.8)*. Then, under* (H_1) *and* (H_2)*, the estimation error satisfies $\|\delta\vartheta\|_F \leqslant r^*$.*

Proof. By convexity of the objective and optimality of $\widehat{\theta}$, each move from θ^* in the direction $t \, \delta\vartheta$ for $t \in [0, 1]$ must lead to a decrease of the objective, *i.e.*

$$
R_n(\theta^* + t \, \delta\vartheta) + \lambda \, |\Omega_{yy}^* + t \, \delta\vartheta_{yy}|_1^- + \mu \, |\Omega_{yx}^* + t \, \delta\vartheta_{yx}|_1 - R_n(\theta^*) - \lambda \, |\Omega_{yy}^*|_1^- - \mu \, |\Omega_{yx}^*|_1 \leqslant 0.
$$

Taking the notation of (5.8), this can be rewritten as $\Delta_n(\theta^* + t \, \delta\vartheta, \theta^*) \leqslant 0$. If $\|\delta\vartheta\|_F \leqslant r^*$ then choose $t = 1$, otherwise calibrate $0 < t < 1$ such that $\|t \, \delta\vartheta\|_F = r^*$. Then, from Lemma

5.9, it clearly follows that $\theta^* + t\,\delta\vartheta \in N_{r,\alpha}(\theta^*)$. Hence, the reasoning preceding (5.8) still holds and, together with Lemma 5.10, we obtain

$$\begin{aligned} 0 &\geqslant \underline{c}\left(|[t\,\delta\vartheta_{yy}]_{\bar{S}}|_1 + |[t\,\delta\vartheta_{yx}]_{\bar{S}}|_1\right) - \overline{c}\left(|[t\,\delta\vartheta_{yy}]_S|_1 + |[t\,\delta\vartheta_{yx}]_S|_1\right) + \gamma_{r,\eta,\beta,p}\,\|t\,\delta\vartheta\|_F^2 \\ &\geqslant -\overline{c}\,|[t\,\delta\vartheta]_S|_1 + \gamma_{r,\eta,\beta,p}\,\|t\,\delta\vartheta\|_F^2 \\ &\geqslant -c_{\lambda,\mu}\max\{h_a,h_b\}\,\sqrt{|S|}\,\|t\,\delta\vartheta\|_F + \gamma_{r,\eta,\beta,p}\,\|t\,\delta\vartheta\|_F^2 \end{aligned}$$

where we used $\underline{c} > 0$ and Cauchy-Schwarz inequality to get $|[\cdot]_S|_1^2 \leqslant |S|\,\|[\cdot]_S\|_F^2$. The constant $c_{\lambda,\mu}$ may be explicitely computed from the configuration of (λ,μ,η) and is given in (A.8). Note that in the proof of Lemma 5.9, it was sufficient to see that $R_n(\theta) - R_n(\theta^*) \geqslant \phi'(0)$ whereas here, we must consider $R_n(\theta) - R_n(\theta^*) = \phi'(0) + e_1(\theta,\theta^*)$ to meet our purposes. That explains the presence of $\gamma_{r,\eta,\beta,p}\,\|t\,\delta\vartheta\|_F^2$ in the inequality. We deduce that the error must satisfy

$$\|t\,\delta\vartheta\|_F \leqslant \frac{c_{\lambda,\mu}\sqrt{|S|}\max\{h_a,h_b\}}{\gamma_{r,\eta,\beta,p}}.$$

As a corollary, it holds that $\|\delta\vartheta\|_F > r^* \;\Rightarrow\; c_{\lambda,\mu}\sqrt{|S|}\max\{h_a,h_b\} \geqslant r^*\,\gamma_{r,\eta,\beta,p}$ or, conversely written, $c_{\lambda,\mu}\sqrt{|S|}\max\{h_a,h_b\} < r^*\,\gamma_{r,\eta,\beta,p} \;\Rightarrow\; \|\delta\vartheta\|_F \leqslant r^*$. □

Lemma 5.12. *Assume that λ, μ and η are chosen according to the configuration of the theorem. Then, under* (H_1)*, there exists absolute constants $b_1 > 0$ and $b_2 > 0$ such that, for any $b_3 \in]0,1[$ and as soon as*

$$n \geqslant \max\left\{b_1\,(q + \lceil s_\alpha\rceil\ln(p+q)),\,\ln(10(p+q)^2) - \ln(b_3)\right\},$$

with probability no less that $1 - \mathrm{e}^{-b_2 n} - b_3$ both the random hypothesis (H_2) *is satisfied and the upper bound*

$$\max\{h_a,h_b\} \leqslant 16\,m^*\sqrt{\frac{\ln(10(p+q)^2) - \ln(b_3)}{n}}$$

holds, where h_a and h_b are given in (2.3) *and* (2.4)*, s_α is defined in* (A.5) *and m^* in* (A.9)*. Hence, one can find a minimal number of observations n_0 such that the theorem holds with high probability as soon as $n > n_0$.*

Proof. All the ingredients of the proof are established in [29]. The authors start by recalling that there exists absolute constants $b_1 > 0$ and $b_2 > 0$ such that hypothesis (H_2) is satisfied with probability no less than $1 - \mathrm{e}^{-b_2 n}$ as soon as $n \geqslant b_1\,(q + \lceil s_\alpha\rceil\ln(p+q))$. We also refer the reader to Lem. 5.1 and Thm. 5.2 of [3], or to Lem. 7.4 of [10] for the random bounds of the restricted isometry constants. Afterwards, they prove (see Prop. 4) that, as soon as $n \geqslant \ln(10(p+q)^2) - \ln(b_3)$ for some $b_3 > 0$, with probability $1 - b_3$,

$$\max\{h_a,h_b\} \leqslant 16\,m^*\sqrt{\frac{\ln(10(p+q)^2) - \ln(b_3)}{n}}.$$

To find the minimal number of observations, we just need to make sure that the above bound is itself smaller than the one of Lemma 5.11. It is then not hard to see that we may retain the minimal size n_0 given in (2.6). □

Appendix A. Some constants

This appendix is entirely dedicated to the constants appearing in the theoretical guarantees. Indeed, a centralization seemed necessary to clarify the rest of the paper, especially the understanding of the main theorem. First, we need to define some constants related to L and to the true values of the model. The bounds

$$\underline{\omega}_L = \frac{\lambda_{\min}(\Omega^*_{yx}\, L\, \Omega^{*\,t}_{yx})}{4\,\lambda_{\max}(\Omega^*_{yy})}, \quad \overline{\omega}_L = \frac{4\,\lambda_{\max}(\Omega^*_{yx}\, L\, \Omega^{*\,t}_{yx})}{\lambda_{\min}(\Omega^*_{yy})}, \quad \overline{\omega}_S = \frac{4\,\lambda_{\max}(\Omega^*_{yx}\, \Sigma^*_{xx}\, \Omega^{*\,t}_{yx})}{\lambda_{\min}(\Omega^*_{yy})}. \tag{A.1}$$

are useful to control the eigenvalues of some recurrent expressions (Lemmas 5.7 and 5.8), uniformly in a neighborhood of $\theta^* = (\Omega^*_{yy}, \Omega^*_{yx})$. The true value of the term at the heart of the structural regularization is

$$s_L = \langle\!\langle L, \Omega^{*\,t}_{yx}\, \Omega^{*\,-1}_{yy}\, \Omega^*_{yx} \rangle\!\rangle. \tag{A.2}$$

It plays a role in the proof of Lemma 5.9 and, as a consequence, in the definition of the area of validity Λ. This important lemma also requires to define

$$\ell_a = |\Omega^{*\,-1}_{yy}\, \Omega^*_{yx}\, L\, \Omega^{*\,t}_{yx}\, \Omega^{*\,-1}_{yy}|_\infty \quad \text{and} \quad \ell_b = 2\,|\Omega^{*\,-1}_{yy}\, \Omega^*_{yx}\, L|_\infty \tag{A.3}$$

and, in the context of the theorem,

$$\alpha = \frac{\max\left\{\frac{(c_\lambda+1)\lambda}{c_\lambda} + \eta\beta\, s_L^{\beta-1}\ell_a, \frac{(c_\mu+1)\mu}{c_\mu} + \eta\beta\, s_L^{\beta-1}\ell_b\right\}}{\min\left\{\frac{(c_\lambda-1)\lambda}{c_\lambda} - \eta\beta\, s_L^{\beta-1}\ell_a, \frac{(c_\mu-1)\mu}{c_\mu} - \eta\beta\, s_L^{\beta-1}\ell_b\right\}}. \tag{A.4}$$

From α and the cardinality of the true active set $|S|$, let

$$s_\alpha = |S|\,\left[1 + \frac{12\,\alpha^2\,\lambda_{\max}(\Sigma^*_{xx})}{\lambda_{\min}(\Sigma^*_{xx})}\right] \tag{A.5}$$

which serves as an upper bound in the random hypothesis (H_2). Similarly, let

$$r^* = \min\{r^*_1, r^*_2, r^*_3, r^*_4\} \tag{A.6}$$

where

$$r^*_1 = \frac{\lambda_{\min}(\Omega^*_{yy})}{2}, \quad r^*_2 = \frac{\frac{\sqrt{10}-\sqrt{7}}{\sqrt{5}}\sqrt{\lambda_{\max}(\Omega^*_{yx}\,\Sigma^*_{xx}\,\Omega^{*\,t}_{yx})}}{\frac{3\sqrt{3}}{2\sqrt{2}}\sqrt{\lambda_{\max}(\Sigma^*_{xx})}}, \quad r^*_3 = \frac{\lambda_{\min}(\Omega^*_{yx}\, L\, \Omega^{*\,t}_{yx})}{4\,\|L\,\Omega^{*\,t}_{yx}\|_2}$$

and

$$r^*_4 = \frac{(\sqrt{2}-1)\sqrt{\lambda_{\max}(\Omega^*_{yx}\, L\, \Omega^{*\,t}_{yx})}}{\sqrt{\lambda_{\max}(L)}}.$$

Together with α given above, r^* is necessary to build the so-called neighborhood $N_{r,\alpha}(\theta^*)$ defined in (5.7), which plays a fundamental role in all our reasonings. It is important to note that, under the configuration of the theorem and hypothesis (H_1), $\alpha > 0$ and $r^* > 0$. Then, Lemma 5.10 highlights a new constant, characterizing a strong local convexity of the smooth part of the objective in the neighborhood $N_{r,\alpha}(\theta^*)$,

$$\gamma_{r,\eta,\beta,p} = \min\left\{\frac{a_1}{8\,\lambda^2_{\max}(\Omega^*_{yy})}, \frac{a_2\,\lambda_{\min}(L)}{4\,\lambda_{\max}(\Omega^*_{yy})} + \frac{a_3\,\lambda_{\min}(\Sigma^*_{xx})}{40\,\lambda_{\max}(\Omega^*_{yy})}\right\} \tag{A.7}$$

where, as it is detailed in the proof of the lemma in question,

$$a_1 = 1 - \epsilon_S\,\overline{\omega}_S - \eta\beta\,p^{\beta-1}\,\overline{\omega}_L^{\beta}\,\epsilon_L, \quad a_2 = \frac{2\,\epsilon_S}{2+\epsilon_S} \quad \text{and} \quad a_3 = \eta\beta\,(p\,\underline{\omega}_L)^{\beta-1}\frac{2\,\epsilon_L}{2+\epsilon_L}$$

for some well-chosen $\epsilon_S > 0$ and $\epsilon_L > 0$. Here again, we make sure that $\gamma_{r,\eta,\beta,p} > 0$. In the same way, in the context of the theorem,

$$c_{\lambda,\mu} = \max\left\{\frac{(c_\lambda+1)\,d_\lambda}{c_\lambda} + e_\lambda,\ \frac{(c_\mu+1)\,d_\mu}{c_\mu} + e_\mu\right\} \tag{A.8}$$

is needed through Lemma 5.11. Finally, independently of the structure matrix L,

$$m^* = |\mathrm{diag}(\Sigma^*_{xx})|_\infty + |\mathrm{diag}(\Omega^{*\,-1}_{yy}\,\Omega^*_{yx}\,\Sigma^*_{xx}\,\Omega^{*\,t}_{yx}\,\Omega^{*\,-1}_{yy})|_\infty \tag{A.9}$$

is going to play a significative role in the upper bound of the theorem.

References

[1] Andrew, G., and Gao, J. Scalable training of L_1-regularized log-linear models. *Proc. 24th Inte. Conf. Mach. Learning.* (2007), 33–40.

[2] Banerjee, O., El Ghaoui, L., and d'Aspremont, A. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res. 9* (2008), 485–516.

[3] Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M. A simple proof of the restricted isometry property for random matrices. *Constr. Approx. 28*, 3 (2008), 253–263.

[4] Boyd, S., and Vandenberghe, L. *Convex Optimization*. Cambridge University Press. 2004.

[5] Cai, T., Li, H., Liu, W., and Xie, J. Covariate-adjusted precision matrix estimation with an application in genetical genomics. *Biometrika. 100*, 1 (2013), 139–156.

[6] Cai, T., Liu, W., and Luo, X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc. 106*, 494 (2011), 594–607.

[7] Chiquet, J., Mary-Huard, T., and Robin, S. Structured regularization for conditional Gaussian graphical models. *Stat. Comput. 27*, 3 (2017), 789–804.

[8] Fan, J., Feng, Y., and Wu, Y. Network exploration via the adaptive Lasso and SCAD penalties. *Ann. Appl. Stat. 3*, 2 (2009), 521–541.

[9] Friedman, J., Hastie, T., and Tibshirani, R. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics. 9*, 3 (2008), 432–441.

[10] Giraud, C. *Introduction to High-Dimensional Statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014.

[11] Hastie, T., Tibshirani, R., and Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC Press, 2015.

[12] Horn, R. A., and Johnson, C. R. *Matrix Analysis (Second Edition)*. Cambridge University Press, Cambridge, New-York, 2012.

[13] Johnson, C., Jalali, A., and Ravikumar, P. High-dimensional sparse inverse covariance estimation using greedy methods. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics.* (2012), vol. 22 of *Proceedings of Machine Learning Research*, PMLR, pp. 574–582.

[14] Lee, W., and Liu, Y. Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood. *J. Multivariate. Anal. 111* (2012), 241–255.

[15] Lu, Z. Smooth optimization approach for sparse covariance selection. *Siam. J. Optimiz. 19*, 4 (2009), 1807–1827.

[16] Maathuis, M., Drton, M., Lauritzen, S. L., and Wainwright, M. *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2018.

[17] Meinshausen, N., and Bühlmann, P. High-dimensional graphs and variable selection with the Lasso. *Ann. Stat. 34*, 3 (2006), 1436–1462.

[18] Pascal, F., Bombrun, L., Tourneret, J. Y., and Berthoumieu, Y. Parameter estimation for multivariate generalized Gaussian distributions. *IEEE. T. Signal. Process. 61*, 23 (2013), 5960–5971.

[19] Peng, J., Wang, P., Zhou, N., and Zhu, J. Partial correlation estimation by joint sparse regression models. *J. Am. Stat. Assoc. 104*, 486 (2009), 735–746.
[20] Ramsay, J., and Silverman, B. *Functional Data Analysis, 2nd ed.* Springer, New-York, 2006.
[21] Ravikumar, P., Wainwright, M., Raskutti, G., and Yu, B. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat. 5* (2011), 935–980.
[22] Rossi, P., Allenby, G., and McCulloch, R. *Bayesian Statistics and Marketing.* Wiley Series in Probability and Statistics. Wiley, 2012.
[23] Rothman, A., Levina, E., and Zhu, J. Sparse multivariate regression with covariance estimation. *J. Comp. Graph. Stat. 19*, 4 (2010), 947–962.
[24] Slawski, M. The structured elastic net for quantile regression and support vector classification. *Stat. Comput. 22* (2012), 153–168.
[25] Slawski, M., Zu Castell, W., and Tutz, G. Feature selection guided by structural information. *Ann. Appl. Stat. 4*, 2 (2010), 1056–1080.
[26] Sohn, K. A., and Kim, S. Joint estimation of structured sparsity and output structure in multiple-output regression via inverse-covariance regularization. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics.* (2012), vol. 22 of *Proceedings of Machine Learning Research*, PMLR, pp. 1081–1089.
[27] Yin, J., and Li, H. A sparse conditional Gaussian graphical model for analysis of genetical genomics data. *Ann. Appl. Stat. 5*, 4 (2011), 2630–2650.
[28] Yuan, M., and Lin, Y. Model selection and estimation in the Gaussian graphical model. *Biometrika. 94*, 1 (2007), 19–35.
[29] Yuan, X. T., and Zhang, T. Partial Gaussian graphical model estimation. *IEEE. T. Inform. Theory. 60*, 3 (2014), 1673–1687.

Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France.
Email address: okome@math.univ-angers.fr

1 Unité de Bioinfomique, Institut de Cancérologie de l'Ouest, Bd Jacques Monod, 44805 Saint Herblain Cedex, France.
2 SIRIC ILIAD, Nantes, Angers, France.
3 CRCINA, INSERM, CNRS, Université de Nantes, Université d'Angers, Institut de Recherche en Santé-Université de Nantes, 8 Quai Moncousu - BP 70721, 44007, Nantes Cedex 1, France.
Email address: pascal.jezequel@ico.unicancer.fr

Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France.
Email address: frederic.proia@univ-angers.fr

2.2 Approche bayésienne

L'estimation de Ω dans un modèle graphique gaussien par inférence bayésienne est une thématique déjà largement développée, elle fait l'objet par exemple de (Maathuis *et al.*, 2018, Chap. 10) où des *a priori* de type Wishart sont étudiés. En revanche, le PGGM bayésien n'avait à notre connaissance jamais été envisagé et c'est pourquoi nous avons souhaité proposer une contrepartie à l'approche fréquentiste. Nos modèles hiérarchiques reposent sur les régressions linéaires de Xu et Ghosh (2015) pour $q = 1$, et sur celles de Liquet *et al.* (2017) généralisées à $q \geqslant 1$, c'est en particulier la raison pour laquelle nous imposons comme ces auteurs la sparsité par une stratégie *spike-and-slab*. Rappelons que nous entendons par là une distribution de probabilité de la forme $(1 - \pi)\,\delta + \pi\,\delta_0$, en d'autres termes une variable aléatoire dont la loi s'exprime ainsi a une probabilité π d'être nulle et une probabilité $1 - \pi$ d'être distribuée selon la loi δ. Un paramètre dont la loi *a posteriori* conserverait cette forme pourrait donc être exactement nul, d'où l'intérêt de la méthode pour engendrer de la sparsité dans une estimation bayésienne. L'article Okome Obiang *et al.* (2022) présenté ci-dessous est accepté pour publication chez *Bayesian Analysis*. Pour la suite et fin de cette section, nous renvoyons le lecteur aux Définitions 1.1–1.4 de l'article en question afin d'avoir plus de détails sur les distributions rencontrées (notations, densités, etc.), certaines étant usuelles mais d'autres un peu moins.

Résumé

Considérons donc un modèle linéaire à réponses multivariées de la forme

$$\mathbb{Y} = -\mathbb{X}\,\Delta^T\Omega_y^{-1} + E \tag{2.12}$$

où $\mathbb{Y} \in \mathbb{R}^{n\times q}$ contient les n réponses, $\mathbb{X} \in \mathbb{R}^{n\times p}$ contient les n prédicteurs, $E \in \mathbb{R}^{n\times q}$ est un bruit multivarié de loi $\mathcal{MN}_{n\times q}(0, I_n, \Omega_y^{-1})$ et le produit $-\Delta^T\Omega_y^{-1} \in \mathbb{R}^{p\times q}$ est la matrice des coefficients de la régression linéaire selon la décomposition évoquée dans la section introductive. Notre procédure a pour principal objectif d'imposer de la sparsité dans Δ, la matrice dont les coefficients sont proportionnels aux corrélations partielles entre les prédicteurs et les réponses. L'effet du i-ème prédicteur sur les réponses se lit dans la i-ème colonne de Δ, on va donc associer la notion de sparsité à celle de colonnes nulles dans cette matrice : $\Delta_i = 0$ revient à dire qu'il n'existe pas de lien direct entre le i-ème prédicteur et les réponses. À cet égard et dans un contexte de grande dimension, on considère trois types de sparsité, selon la terminologie de (Giraud, 2014, Sec. 2.1) :

- *Sparse* (s) où seules certaines colonnes isolées de Δ sont non-nulles.
- *Group-sparse* (gs) où seuls certains groupes de colonnes de Δ sont non-nuls.
- *Sparse-group-sparse* (sgs) où seuls certains groupes de colonnes de Δ sont non-nuls en plus d'être sparse.

Le grand intérêt de cette dernière configuration est qu'elle donne lieu à une sélection à double échelle (groupes et coordonnées). Appelons m le nombre de groupes, κ_g la taille du groupe g et $\underset{\sim}{\Delta}_g$ la sous-matrice de Δ correspondant au groupe g ($1 \leqslant g \leqslant m$). En vertu des relations (23) liant le couple (Ω_y, Δ) aux paramètres de la régression linéaire et de l'inférence bayésienne issue de la littérature (en particulier les deux articles précités),

on propose le modèle hiérarchique

$$\begin{cases} \mathbb{Y} \,|\, \mathbb{X}, \Delta, \Omega_y & \sim & \mathcal{MN}_{n\times q}(-\mathbb{X}\,\Delta^T\,\Omega_y^{-1},\, I_n,\, \Omega_y^{-1}) \\ \underline{\Delta}_g \,|\, \nu_g, \lambda_g, \pi & \overset{\perp\!\!\!\perp}{\sim} & (1-\pi_1)\left[(1-\pi_2)\,\mathcal{N}_q(0,\, \lambda_g\,\nu_{gi}\,\Omega_y) + \pi_2\,\delta_0\right]^{\otimes \kappa_g} + \pi_1\,\delta_0 \\ \nu_{gi} & \overset{\perp\!\!\!\perp}{\sim} & \Gamma(\alpha,\, \ell_{gi}) \\ \lambda_g & \overset{\perp\!\!\!\perp}{\sim} & \Gamma(\alpha_g,\, \gamma_g) \\ \Omega_y & \sim & \mathcal{W}_q(u,\, V) \\ \pi_j & \overset{\perp\!\!\!\perp}{\sim} & \beta(a_j,\, b_j) \end{cases} \tag{2.13}$$

pour $g \in [\![1, m]\!]$, $i \in [\![1, \kappa_g]\!]$ et $j \in [\![1, 2]\!]$. Les hyperparamètres sont spécifiquement choisis et discutés. Une sélection à double échelle s'opère sur Δ à travers $\underline{\Delta}_g$ qui peut être soit annulé entièrement, soit maintenu dans le modèle mais avec des coordonnées annulées. Cette écriture (sgs) se veut très générale mais elle engendre nombre de cas d'intérêt. Ainsi pour $\pi_1 = 0$ et $\lambda_g = 1$, on obtient (s) alors que pour $\pi_2 = 0$ et $\nu_{gi} = 1$, on retrouve (gs). De même, en fixant $\ell_{gi} = \ell_g$ ou $\ell_{gi} = \ell$, voire même $\gamma_g = \gamma$, on considère des situations de *shrinkage* local, par groupe ou global. Avec $\pi_1 = \pi_2 = 0$, on génère un modèle sans sparsité qui pourrait correspondre à la version bayésienne non-pénalisée du PGGM. On tire donc de nombreuses configurations particulières à partir de cette formulation. Sous des hypothèses très spécifiques et suivant le raisonnement de Yang et Narisetty (2020) valable en régression, on montre que dans les cas (s) et (gs), conditionnellement à la variance des réponses,

$$\mathbb{P}(\mathcal{T} \,|\, \mathbb{Y}, \mathbb{X}, \Omega_y) \xrightarrow{\mathbb{P}} 1 \tag{2.14}$$

lorsque $n \to +\infty$, où $\mathcal{T} = \{$le support du vrai modèle est retrouvé$\}$. Des échantillonneurs de Gibbs sont développés afin d'estimer la densité jointe qui découle de ces modèles hiérarchiques et obtenir des estimations *a posteriori*. On a retenu la moyenne *a posteriori* pour Ω_y alors que la médiane *a posteriori* était le choix indiqué pour Δ en raison de sa capacité à fournir des valeurs exactement nulles. Ces méthodes sont ensuite testées en simulations pour comparaison avec les approches contemporaines d'estimation sparse de matrices de précision, ainsi que sur un jeu de données réelles. On pourra trouver les échantillonneurs, les programmes de démonstration ainsi que le jeu de données réelles sur le GitHub `https://github.com/FredericProia/BayesPGGM`.

Perspectives

Les conclusions de cette étude sont très encourageantes, surtout en ce qui concerne la capacité des algorithmes à retrouver le support de Δ. À ce stade les échantillonneurs sont largement améliorables, citons par exemple deux pistes qui devraient absolument être creusées. D'une part, certains termes sont 'dangereux' d'un point de vue numérique car s'écrivant comme produit d'un terme potentiellement très grand et d'un terme potentiellement très petit (typiquement, le terme $|I_{\kappa_g} + \lambda_g\,\underline{\mathbb{X}}_g^T\,\underline{\mathbb{X}}_g|$ que l'on retrouve en dénominateur dans la loi *a posteriori* de $\underline{\Delta}_g$ peut exploser quand κ_g est grand et $\lambda_g > 1$ bien que théoriquement retenu par un terme dont la petitesse équivaut à sa croissance, c'est pourquoi λ_g doit être attentivement contrôlé *via* un choix pertinent de ℓ_g), et la compensation attendue en théorie peut ne pas se produire en pratique. Cela donne lieu à des heuristiques de contrôle des valeurs initiales et des hyperparamètres largement

critiquables. D'autre part, il n'existe pas à notre connaissance d'algorithme simple de simulation de la loi $\mathcal{MGIG}_q$ lorsque $q > 1$. Nous discutons dans l'article d'une approche prometteuse reposant sur la propriété de Matsumoto-Yor (Massam et Wesołowski, 2006, Thm. 3.1), hélas en l'état inapplicable à notre contexte particulier (il faudrait trouver un $z \in \mathbb{R}^q$ tel que $\mathbb{Y}^T\mathbb{Y} + V^{-1} = bzz^T$ pour un $b > 0$, alors que $\mathbb{Y}^T\mathbb{Y} + V^{-1}$ est de rang plein), et cela nous conduit à considérer un échantillonnage par le mode de la loi considérée qui, lui, est facilement accessible par résolution d'une équation de Riccati (cf. Rem. 6.1). En introduisant dans nos simulations un 'oracle' dans lequel les paramètres associés (Ω_y, λ, ν) sont connus, on constate que l'erreur engendrée n'est pas si grande et que cette solution de repli reste satisfaisante en pratique. Mais là encore, cela demanderait à être amélioré lorsque de meilleurs outils de simulation seront à disposition. Par ailleurs, la garantie théorique valable pour (s) et (gs) gagnerait à être étendue à (sgs). Cela ne semble pas insurmontable, mais au prix d'une révision des techniques de preuve de Yang et Narisetty (2020) bien plus conséquente que celle que nous proposons ici. Pour conclure et tenter de faire le lien avec le premier axe de ce mémoire dans lequel nous avons beaucoup insisté sur la problématique de la racine unitaire, on pourrait envisager (déjà pour le cas simplifié de l'AR(1)) un modèle hiérarchique bayésien sur un autorégressif gaussien $\mathbb{Y} = (Y_1, \ldots, Y_n)$ dans lequel on munirait le paramètre d'un *a priori* de type *spike-and-slab* donnant une probabilité $\pi = \pi^+ + \pi^-$ (π^+ pour $+1$, π^- pour -1) à la racine unitaire et une probabilité $1 - \pi$ à la zone de stabilité, de la forme

$$\left\{\begin{array}{lcl} \mathbb{Y} \,|\, \theta, \sigma^2 & \sim & \mathcal{N}_n(0,\, \Sigma)\mathbb{1}_{\{|\theta|<1\}} + \mathcal{N}_n(0,\, \Sigma^+)\mathbb{1}_{\{\theta=+1\}} + \mathcal{N}_n(0,\, \Sigma^-)\mathbb{1}_{\{\theta=-1\}} \\ \theta \,|\, \pi^+, \pi^- & \sim & (1-\pi^+-\pi^-)\,\mathcal{L}(-1,1) + \pi^+\,\delta_{+1} + \pi^-\,\delta_{-1} \\ \sigma^2 & \sim & \mathcal{I}\Gamma(u, v) \\ \pi^+ & \sim & \beta(a^+,\, b^+) \\ \pi^- & \sim & \beta(a^-,\, b^-) \end{array}\right.$$

où Σ, Σ^+ et Σ^- sont les matrices de covariance du vecteur gaussien $\mathbb{Y}$ calculables à partir de θ et σ^2, dépendant du fait que $|\theta| < 1$, $\theta = +1$ ou $\theta = -1$, et où $\mathcal{L}(-1,1)$ est une distribution quelconque portée par $]-1,1[$. Considérant les très bonnes performances en recherche de support des algorithmes que l'on vient de présenter, cette piste pourrait bien conduire à des tests de racines unitaires plus puissants que ceux issus des approches fréquentistes usuelles, c'est pourquoi elle m'intéresse tout particulièrement et des travaux sont en cours. Le passage à l'AR(p) ne sera pas triviale...

A BAYESIAN APPROACH FOR PARTIAL GAUSSIAN GRAPHICAL MODELS WITH SPARSITY

EUNICE OKOME OBIANG, PASCAL JÉZÉQUEL, AND FRÉDÉRIC PROÏA

Abstract. We explore various Bayesian approaches to estimate partial Gaussian graphical models. Our hierarchical structures enable to deal with single-output as well as multiple-output linear regressions, in small or high dimension, enforcing either no sparsity, sparsity, group sparsity or even sparse-group sparsity for a bi-level selection through partial correlations (direct links) between predictors and responses, thanks to spike-and-slab priors corresponding to each setting. Adaptative and global shrinkages are also incorporated in the Bayesian modeling of the direct links. An existing result for model selection consistency is reformulated to stick to our sparse and group-sparse settings, providing a theoretical guarantee under some technical assumptions. Gibbs samplers are developed and a simulation study shows the efficiency of our models which give very competitive results, especially in terms of support recovery. To conclude, a real dataset is investigated.

AMS 2020 subject classifications: Primary 62A09, 62F15; Secondary 62J05.

1. Introduction and Motivations

This paper is devoted to the Bayesian estimation of the partial Gaussian graphical models. Graphical models are now widespread in many contexts, like image analysis, economics or biological regulation networks, neural models, etc. A graphical model for the d-dimensional Gaussian vector $Z \sim \mathcal{N}_d(\mu, \Sigma)$ is a model where the conditional dependencies between the coordinates of Z are represented by means of a graph. We refer the reader to the handbook recently edited by Maathuis *et al.* [18] for a very complete survey of graphical models theory, or to Chap. 7 of Giraud [12] for a wide introduction to the subject. It is well-known that the partial correlation between Z_i and Z_j satisfies

$$\mathbb{C}\mathrm{orr}(Z_i,\, Z_j \,|\, Z_{\neq i,\,j}) = -\frac{\Omega_{ij}}{\sqrt{\Omega_{ii}\,\Omega_{jj}}}$$

where $\Omega = \Sigma^{-1} \in \mathbb{S}^d_{++}$ is the precision matrix of Z (the notation $\mathbb{S}^d_{++}$ for the cone of symmetric positive definite matrices of dimension d is used in all the paper). A fundamental consequence of this is that there is a partial correlation between Z_i and Z_j if and only if the (i, j)-th element of Ω is non-zero. The sparse estimation of Ω is therefore a major issue for variable selection in high-dimensional studies, which has given rise to a substantial literature, see *e.g.* the seminal work of Meinshausen and Bühlmann [20]. This logically led numerous authors to investigate interesting properties under various kind of hypotheses, estimation procedures and penalties. Let us mention for example the optimality results obtained by Cai and Zhou [5] and the penalized estimations of Yuan and Lin [32], Rothman *et al.* [26], Banerjee *et al.* [2], Cai *et al.* [4] or Ravikumar *et al.* [24], all coming with

Key words and phrases. High-dimensional linear regression, Partial graphical model, Partial correlation, Bayesian approach, Sparsity, Spike-and-slab, Gibbs sampler.

theoretical guarantees, algorithmic considerations and real world examples. Besides, the famous graphical Lasso of Friedman *et al.* [10] has become an essential tool for dealing with precision matrix estimation. Perhaps more attractive to us since focusing on each entry of the precision matrix (no longer taken as a whole), the approach of Ren *et al.* [25] is remarkable and will serve as a basis for comparison in our simulation study. The Bayesian inference counterpart has been developed as well, it is *e.g.* the subject of Chap. 10 of Maathuis *et al.* [18] where various Wishart-type priors are considered for Ω, see also Li *et al.* [15] or Gan *et al.* [11] for spike-and-slab approaches and all references within.

Suppose now that we deal with a multivariate linear regression of the form

$$\mathbb{Y} = \mathbb{X}B + E$$

where $\mathbb{Y} \in \mathbb{R}^{n\times q}$ is a matrix of q-dimensional responses of which k-th row is Y_k^t, $\mathbb{X} \in \mathbb{R}^{n\times p}$ is a matrix of p-dimensional predictors of which k-th row is X_k^t, $B \in \mathbb{R}^{p\times q}$ contains the regression coefficients and $E \in \mathbb{R}^{n\times q}$ is a matrix-variate Gaussian noise. The Partial Gaussian Graphical Model (PGGM), developped *e.g.* by Sohn and Kim [27] or Yuan and Zhang [33], appears as a powerful tool to exhibit relations between predictors and responses that exist through partial correlations (called from now on 'direct links', as opposed to 'indirect links' resulting from correlations). Indeed, assume that the couple $(Y_k, X_k) \in \mathbb{R}^{q+p}$ is jointly normally distributed with zero mean, covariance Σ and precision Ω. Then, the block decomposition given by

$$\Omega = \begin{pmatrix} \Omega_y & \Delta \\ \Delta^t & \Omega_x \end{pmatrix}$$

with $\Omega_y \in \mathbb{S}_{++}^q$, $\Delta \in \mathbb{R}^{q\times p}$ and $\Omega_x \in \mathbb{S}_{++}^p$ leads to $Y_k \,|\, X_k \sim \mathcal{N}_q(-\Omega_y^{-1}\Delta\, X_k,\, \Omega_y^{-1})$. This is a crucial remark because one can see that the multiple-output regression $Y_k = B^t\, X_k + E_k$ with Gaussian noise $E_k \sim \mathcal{N}_q(0, R)$ may be reparametrized with

$$B = -\Delta^t\, \Omega_y^{-1} \quad \text{and} \quad R = \Omega_y^{-1}. \tag{1.1}$$

A large volume of literature exists for the sparse estimation of B with arbitrary group structures (see *e.g.* Li *et al.* [14] or Chap. 6 of Giraud [12]), but we will not tackle this issue in our study. At least not frontally but indirectly, since the latter relations show that an estimation of B is possible through the one of the pair (Ω_y, Δ). Whereas B contains direct and indirect links between the predictors and the responses (due *e.g.* to strong correlations among the variables), Δ is clearly more interesting from an inferential point of view for it only contains direct links. However, while the estimation of (Ω_y, Δ) appears to be essential, it usually depends on the accuracy of the estimation of the whole precision matrix, which, in turn, may be strongly affected by the one of Ω_x. For example, the graphical Lasso of Friedman *et al.* [10] involves maximizing the log-likelihood penalized by the elementwise ℓ_1 norm of Ω. For multiple-output high-dimensional regressions where generally $p \gg q$, we understand that a significant bias is likely to result from the large-scale shrinkage. Another substantial advantage of the partial model is that we can override this issue by computing a new objective function in which Ω_x has disappeared, *i.e.* the penalized log-likelihood

$$\begin{aligned} L_n(\Omega_y, \Delta) \;=\; & -\ln\det(\Omega_y) + \operatorname{tr}(S_y\,\Omega_y) + 2\operatorname{tr}(S_{yx}^t\,\Delta) \\ & + \operatorname{tr}(S_x\,\Delta^t\,\Omega_y^{-1}\,\Delta) + \lambda\,\mathrm{pen}(\Omega_y) + \mu\,\mathrm{pen}(\Delta) \end{aligned} \tag{1.2}$$

where $S_x \in \mathbb{S}_{++}^p$ and $S_y \in \mathbb{S}_{++}^q$ are the empirical variances of the responses and the predictors, respectively, and where $S_{yx} \in \mathbb{R}^{q\times p}$ is the empirical covariance, computed on the basis of

a set of n observations. This can be obtained either by considering the multiple-output Gaussian regression scheme, or, as it is done by Yuan and Zhang [33], by eliminating Ω_x thanks to a first optimization step in the objective function of the graphical model. The usual convex penalties are elementwise ℓ_1 norms, possibly deprived of the diagonal terms for Ω_y. This paved the way to the recent study of Chiquet *et al.* [6] where the authors replace the penalty on Ω_y by a structuring one enforcing various kind of sparsity patterns in Δ, and to the one of Okome Obiang *et al.* [21] in which some theoretical guarantees are provided for a slightly more general estimation procedure.

However, to the best of our knowledge, the Bayesian approach for the PGGM is a new research topic. Given the outputs gathered in $\mathbb{Y}$ and the predictors gathered in $\mathbb{X}$, the objective of this paper is the Bayesian estimation of the direct links and the precision matrix of the responses. This is inspired by the ideas of Xu and Ghosh [29] for the single-output setting ($q = 1$), and by the ones of Liquet *et al.* [17] for the multiple-output setting ($q > 1$). Taking advantage of the relations (1.1), we consider that a Gaussian prior for B must remain Gaussian for Δ (with a correctly updated variance), and that an inverse Wishart prior for R merely becomes a Wishart one for Ω_y. Yet, despite these seemingly small changes in the design of the priors, we will see that the resulting distributions are completely different. The hierarchical models that we are going to study all come from this working base, but let us point out that a wide variety of refinements exists in the recent literature for Bayesian sparsity, like the grouped 'horseshoe' of Xu *et al.* [30], the 'aggressive' multivariate Dirichlet-Laplace prior of Wei *et al.* [28], the theoretical results for group selection consistency of Yang and Narisetty [31] or even the extension of the Bayesian spike-and-slab group selection to generalized additive models of Bai *et al.* [1], all related to the regression setting but that might also be investigated for PGGMs. To enforce various types of sparsity in Δ for high-dimensional problems, we decided to make use of spike-and-slab priors, with a spike probability guided by a conjugate Beta distribution.

The paper is organized as follows. Sections 2, 3 and 4 are dedicated to the study of our hierarchical models enforcing either no sparsity, sparsity, group sparsity or sparse-group sparsity in the direct links, respectively, according to the terminology of Sec. 2.1 of Giraud [12]. In particular, we will see that our bi-level selection clearly diverges from the strategy of Liquet *et al.* [17]. We also adapt the reasoning of Yang and Narisetty [31] to establish group selection consistency under some technical assumptions and an appropriate amount of sparsity. Section 5 is devoted to the conditional posterior distributions of the parameters in order to implement Gibbs samplers that are tested in Section 6. This empirical section is focused on a simulation study first, to evaluate and compare the efficiency of the models, then a real dataset is treated, and a short conclusion ends the paper. But, firstly, let us give some examples of what exactly we mean by 'sparse', 'group-sparse' and 'sparse-group-sparse' settings, and let us summarize the definitions that we have chosen to retain for the well-known distributions as well as for the less usual ones, in order to avoid any misinterpretation of our results and proofs.

Example 1.1. To explain a set of phenotypic traits, suppose that we investigate a large collection of genetic markers spread over twenty chromosomes. For coordinate sparsity ('sparse' setting), only a few markers are active. For group sparsity ('group-sparse' setting), the markers are clustered into groups (formed by chromosomes) and only a few of them are active. For sparse-group sparsity ('sparse-group-sparse' setting), only a few chromosomes are active

and they are sparse, the result is a bi-level selection (chromosomes and markers). This will be the context of our example on real data (Section 6.2).

Definition 1.1 (Gaussian)**.** *The density of $X \in \mathbb{R}^{d_1 \times d_2}$ following the matrix normal distribution $\mathcal{MN}_{d_1 \times d_2}(M,\, \Sigma_1,\, \Sigma_2)$ is given by*

$$p(X) = \frac{1}{(2\pi)^{\frac{d_1\, d_2}{2}}\, |\Sigma_1|^{\frac{d_2}{2}}\, |\Sigma_2|^{\frac{d_1}{2}}}\, \exp\!\left(-\frac{1}{2}\,\mathrm{tr}\big(\Sigma_2^{-1}(X-M)^t\,\Sigma_1^{-1}(X-M)\big)\right)$$

where $M \in \mathbb{R}^{d_1 \times d_2}$, $\Sigma_1 \in \mathbb{S}_{++}^{d_1}$ and $\Sigma_2 \in \mathbb{S}_{++}^{d_2}$. When $d_2 = 1$, this is a multivariate normal distribution $\mathcal{N}_d(\mu,\, \Sigma)$ with $d = d_1$, $\mu = M$ and $\Sigma = \Sigma_2^{-1}\Sigma_1$, having density

$$p(X) = \frac{1}{(2\pi)^{\frac{d}{2}}\, |\Sigma|^{\frac{1}{2}}}\, \exp\!\left(-\frac{1}{2}\,(X-\mu)^t\,\Sigma^{-1}(X-\mu)\right)$$

where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{S}_{++}^d$.

Definition 1.2 (Generalized Inverse Gaussian)**.** *The density of $X \in \mathbb{S}_{++}^d$ following the matrix generalized inverse Gaussian distribution $\mathcal{MGIG}_d(\nu,\, A,\, B)$ is given by*

$$p(X) = \frac{|X|^{\nu - \frac{d+1}{2}}}{\left|\frac{A}{2}\right|^{\nu} B_\nu\big(\frac{A}{2}, \frac{B}{2}\big)}\, \exp\!\left(-\frac{1}{2}\,\mathrm{tr}\big(A\,X^{-1} + B\,X\big)\right) \mathbb{1}_{\{X \in \mathbb{S}_{++}^d\}}$$

where $\nu \in \mathbb{R}$, $A \in \mathbb{S}_{++}^d$, $B \in \mathbb{S}_{++}^d$ and B_ν is a Bessel-type function of order ν. When $d = 1$, this is a generalized inverse Gaussian distribution $\mathcal{GIG}(\nu,\, a,\, b)$ with $a = A$ and $b = B$, having density

$$p(X) = \frac{X^{\nu-1}}{\big(\frac{a}{2}\big)^{\nu} B_\nu\big(\frac{a}{2}, \frac{b}{2}\big)}\, \mathrm{e}^{-\frac{a}{2X} - \frac{b\,X}{2}}\, \mathbb{1}_{\{X > 0\}}$$

where $\nu \in \mathbb{R}$, $a > 0$ and $b > 0$.

Definition 1.3 (Wishart/Gamma/Exponential)**.** *The density of $X \in \mathbb{S}_{++}^d$ following the matrix Wishart distribution $\mathcal{W}_d(u,\, V)$ is given by*

$$p(X) = \frac{|X|^{\frac{u-d-1}{2}}}{2^{\frac{d\,u}{2}}\, \Gamma_d\big(\frac{u}{2}\big)\, |V|^{\frac{u}{2}}}\, \exp\!\left(-\frac{1}{2}\,\mathrm{tr}\big(V^{-1}X\big)\right) \mathbb{1}_{\{X \in \mathbb{S}_{++}^d\}}$$

where $u > d-1$, $V \in \mathbb{S}_{++}^d$ and Γ_d is the multivariate Gamma function of order d. When $d = 1$, this is a Gamma distribution $\Gamma(a,\, b)$ with $a = \frac{u}{2}$ and $\frac{1}{b} = 2\,V$, having density

$$p(X) = \frac{b^a\, X^{a-1}}{\Gamma(a)}\, \mathrm{e}^{-b\,X}\, \mathbb{1}_{\{X > 0\}}$$

where $a > 0$ and $b > 0$. The exponential distribution $\mathcal{E}(\ell)$ is then defined as the $\Gamma(1,\, \ell)$ distribution, for $\ell > 0$.

Definition 1.4 (Beta)**.** *The density of $X \in [0, 1]$ following the Beta distribution $\beta(a,\, b)$ is given by*

$$p(X) = \frac{X^{a-1}\,(1-X)^{b-1}}{\beta(a,b)}\, \mathbb{1}_{\{0 \leqslant X \leqslant 1\}}$$

where $a > 0$, $b > 0$ and β is the Beta function.

In all the paper, data and parameters are gathered in $\Theta = \{\mathbb{Y}, \mathbb{X}, \Delta, \Omega_y, \nu, \lambda, \pi\}$ and, to standardize, for any $e \in \Theta$, we note $\Theta_e = \Theta \backslash \{e\}$.

2. The sparse setting

In this section, $\lambda_i \in \mathbb{R}$ is the i-th component of $\lambda \in \mathbb{R}^p$, $\Delta_i \in \mathbb{R}^q$ is the i-th column of Δ and $\mathbb{X}_i \in \mathbb{R}^n$ stands for the i-th column of $\mathbb{X}$ ($1 \leqslant i \leqslant p$). Let us consider the hierarchical Bayesian model, where the columns of Δ are assumed to be independent, given by

$$\left\{\begin{array}{lcl} \mathbb{Y} \,|\, \mathbb{X}, \Delta, \Omega_y & \sim & \mathcal{MN}_{n\times q}(-\mathbb{X}\,\Delta^t\,\Omega_y^{-1},\, I_n,\, \Omega_y^{-1}) \\ \Delta_i \,|\, \Omega_y, \lambda_i, \pi & \overset{\perp\!\!\!\perp}{\sim} & (1-\pi)\,\mathcal{N}_q(0,\, \lambda_i\,\Omega_y) + \pi\,\delta_0 \\ \lambda_i & \overset{\perp\!\!\!\perp}{\sim} & \Gamma(\alpha,\, \ell_i) \\ \Omega_y & \sim & \mathcal{W}_q(u,\, V) \\ \pi & \sim & \beta(a,\, b) \end{array}\right. \tag{2.1}$$

for $i \in [\![1, p]\!]$, with hyperparameters $\alpha = \frac{1}{2}(q+1)$, $\ell_i > 0$, $u > q-1$, $V \in \mathbb{S}_{++}^q$, $a > 0$ and $b > 0$. A general ungrouped sparsity is promoted in the columns of Δ through the spike-and-slab prior. In this mixture model, π is the prior spike probability and λ is an adaptative shrinkage factor acting at the predictor scale (λ_i is associated with the direct links between predictor i and all the responses). When $\ell_i = \ell$ for all i, we will rather speak of global shrinkage. The degree of sparsity will be characterized by the number N_0 of zero columns of Δ, that is

$$N_0 = \mathrm{Card}(i,\, \Delta_i = 0) = \sum_{i=1}^{p} \mathbb{1}_{\{\Delta_i = 0\}}. \tag{2.2}$$

To implement a Gibbs sampler from the full posterior distribution stemming from (2.1), we may use the conditional distributions given in the proposition below.

Proposition 2.1. *In the hierarchical model* (2.1)*, the conditional posterior distributions are as follows.*

– *The parameter* Δ *satisfies, for* $i \in [\![1, p]\!]$,

$$\Delta_i \,|\, \Theta_{\Delta_i} \;\sim\; (1-p_i)\,\mathcal{N}_q(-s_i\,H_i,\, s_i\,\Omega_y) + p_i\,\delta_0$$

where

$$H_i = \Omega_y\,\mathbb{Y}^t\,\mathbb{X}_i + \sum_{j\neq i}\langle \mathbb{X}_i,\, \mathbb{X}_j\rangle\,\Delta_j, \qquad s_i = \frac{\lambda_i}{1+\lambda_i\,\|\mathbb{X}_i\|^2}$$

and

$$p_i = \frac{\pi}{\pi + (1-\pi)\,(1+\lambda_i\,\|\mathbb{X}_i\|^2)^{-\frac{q}{2}}\,\exp\Big(\frac{s_i\,H_i^t\,\Omega_y^{-1}H_i}{2}\Big)}.$$

– *The parameter* Ω_y *satisfies*

$$\Omega_y \,|\, \Theta_{\Omega_y} \;\sim\; \mathcal{MGIG}_q\bigg(\frac{n-p+N_0+u}{2},\, \Delta\,(\mathbb{X}^t\,\mathbb{X} + D_\lambda^{-1})\,\Delta^t,\, \mathbb{Y}^t\,\mathbb{Y} + V^{-1}\bigg)$$

where $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$.

– *The parameter* λ *satisfies, for* $i \in [\![1, p]\!]$,

$$\lambda_i \,|\, \Theta_{\lambda_i} \;\sim\; \mathbb{1}_{\{\Delta_i \neq 0\}}\,\mathcal{GIG}\bigg(\frac{1}{2},\, \Delta_i^t\,\Omega_y^{-1}\Delta_i,\, 2\,\ell_i\bigg) + \mathbb{1}_{\{\Delta_i = 0\}}\,\Gamma(\alpha,\, \ell_i).$$

– *The parameter* π *satisfies*

$$\pi \,|\, \Theta_\pi \;\sim\; \beta\big(N_0 + a,\, p - N_0 + b\big).$$

Proof. See Section 5.1. □

Remark 2.1. The Bayesian Lasso, as introduced *e.g.* in Sec. 6.1 of [13] or in [22], assumes a prior Laplace distribution for the regression coefficients conditional on the noise variance. In our case, $\Delta_i \,|\, \Omega_y, \pi$ is still a multivariate spike-and-slab (after integrating over λ_i), with a slab following a so-called multivariate K-distribution (see [7]), which is a generalization of the multivariate Laplace distribution. See *e.g.* Sec 2.1 of [17]. From this point of view, our study is in line with the usual Bayesian regression schemes. Perhaps even more interesting, going on with the idea of the authors, suppose that, for all $1 \leqslant i \leqslant p$, $\Delta_i = b_i \, \Delta_i^*$ where Δ_i^* follows the multivariate K-distribution described above and $b_i \,|\, \pi \sim \mathcal{B}(1-\pi)$ is independent of Δ_i^*. Now, the sparsity in Δ is not induced by a spike-and-slab strategy anymore but, equivalently, by multiplying the slab part by an independent Bernoulli variable being 0 with probability π. Then, it is possible to show that the negative log-likelihood of this alternative hierarchical model is given, up to an additive constant that does not depend on Δ, by

$$\frac{1}{2} \left\| (\mathbb{Y} + \mathbb{X}\,\Delta^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}} \right\|_F^2 + \sum_{i=1}^{p} c_i \left\| \Omega_y^{-\frac{1}{2}} \Delta_i^* \right\|_F + \ln\!\left(\frac{1-\pi}{\pi}\right) \sum_{i=1}^{p} b_i$$

where $c_i > 0$. We first recognize an ℓ_2-type penalty but also an ℓ_0-type penalty on Δ (provided that $\pi < \frac{1}{2}$) since summing the b_i amounts to counting the number of non-zero columns in Δ. Consequently, there is a close connection between our hierarchical Bayesian model and the regressions penalized by ℓ_2 and ℓ_0 norms, problems that are known to be very hard to solve due to combinatorial optimization.

The particular case $q = 1$ is a very useful corollary of the proposition. Here, the direct links form a row vector such that $\Delta^t \in \mathbb{R}^p$ with components $\Delta_i \in \mathbb{R}$ $(1 \leqslant i \leqslant p)$, and the precision matrix of the responses reduces to $\omega_y > 0$. According to the parametrization of the distributions (see Section 1), the corresponding prior distribution of ω_y is $\Gamma(\frac{u}{2}, \frac{1}{2\,v})$ for $u, v > 0$ and the one of λ_i is $\mathcal{E}(\ell_i)$ for $\ell_i > 0$. The other priors are unchanged.

Corollary 2.1. *In the hierarchical model* (2.1) *with* $q = 1$*, the conditional posterior distributions are as follows.*

– *The parameter* Δ *satisfies, for* $i \in [\![1, p]\!]$*,*

$$\Delta_i \,|\, \Theta_{\Delta_i} \;\sim\; (1-p_i)\,\mathcal{N}(-s_i\,h_i,\; s_i\,\omega_y) + p_i\,\delta_0$$

where

$$h_i = \omega_y\,\langle \mathbb{X}_i,\, \mathbb{Y} \rangle + \sum_{j \neq i} \langle \mathbb{X}_i,\, \mathbb{X}_j \rangle\,\Delta_j, \qquad s_i = \frac{\lambda_i}{1 + \lambda_i\,\|\mathbb{X}_i\|^2}$$

and

$$p_i = \frac{\pi}{\pi + (1-\pi)\,(1 + \lambda_i\,\|\mathbb{X}_i\|^2)^{-\frac{1}{2}}\,\exp\!\left(\frac{s_i\,h_i^2}{2\,\omega_y}\right)}.$$

– *The parameter* ω_y *satisfies*

$$\omega_y \,|\, \Theta_{\omega_y} \;\sim\; \mathcal{GIG}\!\left(\frac{n - p + N_0 + u}{2},\; \Delta\,(\mathbb{X}^t\,\mathbb{X} + D_\lambda^{-1})\,\Delta^t,\; \|\mathbb{Y}\|^2 + \frac{1}{v}\right)$$

where $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$.

– *The parameter* λ *satisfies, for* $i \in [\![1, p]\!]$,

$$\lambda_i \,|\, \Theta_{\lambda_i} \;\sim\; \mathbb{1}_{\{\Delta_i \neq 0\}}\, \mathcal{GIG}\left(\frac{1}{2},\, \frac{\Delta_i^2}{\omega_y},\, 2\,\ell_i\right) + \mathbb{1}_{\{\Delta_i = 0\}}\, \mathcal{E}(\ell_i).$$

– *The parameter* π *satisfies*

$$\pi \,|\, \Theta_\pi \;\sim\; \beta\big(N_0 + a,\, p - N_0 + b\big).$$

Proof. This is a consequence of Proposition 2.1. □

Note that we can also easily derive the Bayesian counterpart of the standard PGGM adapted to the small-dimensional case, with no sparsity, by taking $\pi = 0$.

Corollary 2.2. *In the hierarchical model* (2.1) *with* $\pi = 0$*, the conditional posterior distributions are as follows.*

– *The parameter* Δ *satisfies, for* $i \in [\![1, p]\!]$,

$$\Delta_i \,|\, \Theta_{\Delta_i} \;\sim\; \mathcal{N}_q(-s_i\, H_i,\, s_i\, \Omega_y)$$

where

$$H_i = \Omega_y\, \mathbb{Y}^t\, \mathbb{X}_i + \sum_{j \neq i} \langle \mathbb{X}_i,\, \mathbb{X}_j \rangle\, \Delta_j \quad \textit{and} \quad s_i = \frac{\lambda_i}{1 + \lambda_i\, \|\mathbb{X}_i\|^2}.$$

– *The parameter* Ω_y *satisfies*

$$\Omega_y \,|\, \Theta_{\Omega_y} \;\sim\; \mathcal{MGIG}_q\left(\frac{n - p + u}{2},\, \Delta\,(\mathbb{X}^t\,\mathbb{X} + D_\lambda^{-1})\,\Delta^t,\, \mathbb{Y}^t\,\mathbb{Y} + V^{-1}\right)$$

where $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$.

– *The parameter* λ *satisfies, for* $i \in [\![1, p]\!]$,

$$\lambda_i \,|\, \Theta_{\lambda_i} \;\sim\; \mathcal{GIG}\left(\frac{1}{2},\, \Delta_i^t\, \Omega_y^{-1} \Delta_i,\, 2\,\ell_i\right).$$

Proof. This is a consequence of Proposition 2.1. □

In the simulation study of Section 6.1, Scen. 0, 1 and 2 are dedicated to the sparse setting. The next section discusses the group sparsity in Δ.

3. The group-sparse setting

The predictors are now ordered in m groups of sizes $\kappa_1 + \ldots + \kappa_m = p$. For the g-th group ($1 \leqslant g \leqslant m$), $\lambda_g \in \mathbb{R}$ is the g-th component of $\lambda \in \mathbb{R}^m$, the covariate submatrix is $\underline{\mathbb{X}}_g \in \mathbb{R}^{n \times \kappa_g}$ and the corresponding slice of Δ is $\underline{\Delta}_g \in \mathbb{R}^{q \times \kappa_g}$. Let us consider the hierarchical Bayesian model, where the columns of Δ are assumed to be independent both within and between the groups, given by

$$\text{(3.1)} \qquad \begin{cases} \mathbb{Y} \,|\, \mathbb{X}, \Delta, \Omega_y & \sim & \mathcal{MN}_{n \times q}(-\mathbb{X}\,\Delta^t\,\Omega_y^{-1},\, I_n,\, \Omega_y^{-1}) \\ \underline{\Delta}_g \,|\, \Omega_y, \lambda_g, \pi & \overset{\perp\!\!\!\perp}{\sim} & (1-\pi)\,\mathcal{MN}_{q \times \kappa_g}(0,\, \lambda_g\,\Omega_y, I_{\kappa_g}) + \pi\,\delta_0 \\ \lambda_g & \overset{\perp\!\!\!\perp}{\sim} & \Gamma(\alpha_g,\, \ell_g) \\ \Omega_y & \sim & \mathcal{W}_q(u,\, V) \\ \pi & \sim & \beta(a,\, b) \end{cases}$$

for $g \in [\![1, m]\!]$, with hyperparameters $\alpha_g = \frac{1}{2}(q\,\kappa_g + 1)$, $\ell_g > 0$, $u > q-1$, $V \in \mathbb{S}^q_{++}$, $a > 0$ and $b > 0$. A general group sparsity is promoted in the columns of Δ through the spike-and-slab prior at the group level. In this mixture model, π is the prior spike probability and λ is an adaptative shrinkage factor acting at the group scale (λ_g is associated with the direct links between the predictors of group g and all the responses). Likewise, when $\ell_g = \ell$ for all g, we will rather speak of global shrinkage. Now, the degree of sparsity will be characterized by N_0 given in (2.2), but also by the number G_0 of zero groups of Δ, that is

$$G_0 = \mathrm{Card}(g,\, \underline{\Delta}_g = 0) = \sum_{g=1}^{m} \mathbb{1}_{\{\underline{\Delta}_g = 0\}}. \tag{3.2}$$

To implement a Gibbs sampler from the full posterior distribution stemming from (3.1), we may use the conditional distributions given in the proposition below.

Proposition 3.1. *In the hierarchical model* (3.1)*, the conditional posterior distributions are as follows.*

– *The parameter* Δ *satisfies, for* $g \in [\![1, m]\!]$*,*

$$\underline{\Delta}_g \,|\, \Theta_{\underline{\Delta}_g} \;\sim\; (1-p_g)\,\mathcal{MN}_{q\times\kappa_g}(-H_g\,S_g,\, \Omega_y,\, S_g) + p_g\,\delta_0$$

where

$$H_g = \Omega_y \mathbb{Y}^t \underline{\mathbb{X}}_g + \sum_{j\neq g} \underline{\Delta}_j\, \underline{\mathbb{X}}_j^t\, \underline{\mathbb{X}}_g, \qquad S_g = \lambda_g \left(I_{\kappa_g} + \lambda_g\, \underline{\mathbb{X}}_g^t\, \underline{\mathbb{X}}_g\right)^{-1}$$

and

$$p_g = \frac{\pi}{\pi + (1-\pi)\,|I_{\kappa_g} + \lambda_g\, \underline{\mathbb{X}}_g^t\, \underline{\mathbb{X}}_g|^{-\frac{q}{2}}\, \exp\!\Big(\frac{\mathrm{tr}(H_g^t\,\Omega_y^{-1} H_g\, S_g)}{2}\Big)}.$$

– *The parameter* Ω_y *satisfies*

$$\Omega_y \,|\, \Theta_{\Omega_y} \;\sim\; \mathcal{MGIG}_q\!\left(\frac{n-p+N_0+u}{2},\, \Delta\,(\mathbb{X}^t\mathbb{X} + D_\lambda^{-1})\,\Delta^t,\, \mathbb{Y}^t\mathbb{Y} + V^{-1}\right)$$

where $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_1, \ldots, \lambda_m, \ldots, \lambda_m)$ *with each* λ_g *duplicated* κ_g *times.*

– *The parameter* λ *satisfies, for* $g \in [\![1, m]\!]$*,*

$$\lambda_g \,|\, \Theta_{\lambda_g} \;\sim\; \mathbb{1}_{\{\underline{\Delta}_g \neq 0\}}\, \mathcal{GIG}\!\left(\frac{1}{2},\, \mathrm{tr}(\underline{\Delta}_g^t\, \Omega_y^{-1} \underline{\Delta}_g),\, 2\,\ell_g\right) + \mathbb{1}_{\{\underline{\Delta}_g = 0\}}\, \Gamma(\alpha_g,\, \ell_g).$$

– *The parameter* π *satisfies*

$$\pi \,|\, \Theta_\pi \;\sim\; \beta\big(G_0 + a,\, m - G_0 + b\big).$$

Proof. See Section 5.2. □

Note that Remark 2.1 still applies to this configuration, after some adjustments (the ℓ_0-like penalty is on the number of non-zero groups). Here again, the particular case $q = 1$ is a very useful corollary. The direct links form a row vector such that $\Delta^t \in \mathbb{R}^p$ with groups $\underline{\Delta}_g^t \in \mathbb{R}^{\kappa_g}$ ($1 \leqslant g \leqslant m$), the precision matrix of the responses reduces to $\omega_y > 0$. According to the parametrization of the distributions (see Section 1), the corresponding prior distribution of ω_y is $\Gamma(\frac{u}{2}, \frac{1}{2\,v})$ for $u, v > 0$, like in the ungrouped setting. The other priors are unchanged.

Corollary 3.1. *In the hierarchical model* (3.1) *with* $q = 1$*, the conditional posterior distributions are as follows.*

– *The parameter* Δ *satisfies, for* $g \in [\![1, m]\!]$,

$$\underline{\Delta}_g^t \,|\, \Theta_{\Delta_g} \;\sim\; (1-p_g)\,\mathcal{N}_{\kappa_g}(-S_g\,H_g,\, \omega_y\,S_g) + p_g\,\delta_0$$

where

$$H_g = \omega_y\,\underline{\mathbb{X}}_g^t\,\mathbb{Y} + \sum_{j\neq g} \underline{\mathbb{X}}_g^t\,\underline{\mathbb{X}}_j\,\underline{\Delta}_j^t, \qquad S_g = \lambda_g\left(I_{\kappa_g} + \lambda_g\,\underline{\mathbb{X}}_g^t\,\underline{\mathbb{X}}_g\right)^{-1}$$

and

$$p_g = \frac{\pi}{\pi + (1-\pi)\,|I_{\kappa_g} + \lambda_g\,\underline{\mathbb{X}}_g^t\,\underline{\mathbb{X}}_g|^{-\frac{1}{2}}\,\exp\!\Big(\frac{H_g^t\,S_g\,H_g}{2\,\omega_y}\Big)}.$$

– *The parameter* ω_y *satisfies*

$$\omega_y \,|\, \Theta_{\omega_y} \;\sim\; \mathcal{GIG}\!\left(\frac{n-p+N_0+u}{2},\, \Delta\,(\mathbb{X}^t\,\mathbb{X} + D_\lambda^{-1})\,\Delta^t,\, \|\mathbb{Y}\|^2 + \frac{1}{v}\right)$$

where $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_1, \ldots, \lambda_m, \ldots, \lambda_m)$ *with each* λ_g *duplicated* κ_g *times.*

– *The parameter* λ *satisfies, for* $g \in [\![1, m]\!]$,

$$\lambda_g \,|\, \Theta_{\lambda_g} \;\sim\; \mathbb{1}_{\{\Delta_g \neq 0\}}\,\mathcal{GIG}\!\left(\frac{1}{2},\, \frac{\|\underline{\Delta}_g\|^2}{\omega_y},\, 2\,\ell_g\right) + \mathbb{1}_{\{\Delta_g = 0\}}\,\Gamma(\alpha_g,\, \ell_g).$$

– *The parameter* π *satisfies*

$$\pi \,|\, \Theta_\pi \;\sim\; \beta\big(G_0 + a,\, m - G_0 + b\big).$$

Proof. This is a consequence of Proposition 3.1. □

In the simulation study of Section 6.1, Scen. 3 and 4 are dedicated to the group-sparse setting. To conclude this section, a theoretical guarantee is provided (given Ω_y and with $\lambda = \lambda_n$ and $\pi = \pi_n$ depending on n). It is possible to obtain a model selection consistency property for this approach when both the number of observations n and the number of groups $m = m_n$ tend to infinity, by adapting the reasoning of [31] dedicated to the linear regression (with $q = 1$). Indeed, when Ω_y is known, Δ reduces to a linear transformation of B. Thus, it is not surprising that a similar result follows under the same kind of hypotheses. In the sequel, we denote by $\mathbb{X}_{(k)} \in \mathbb{R}^{n\times|k|}$ the design matrix of rank r_k corresponding to the submodel indexed by the binary vector $k \in \{0,1\}^m$ having $|k|$ non-zero values ($k_g = 1$ means that the g-th group is included in the model), and by $\Pi_{(k)} \in \mathbb{R}^{n\times n}$ the projection matrix onto the column-space of $\mathbb{X}_{(k)}$. Similarly, Δ restricted to k is $\Delta_{(k)} \in \mathbb{R}^{q\times|k|}$. The true model is called t and $t^{\pm g}$ are submodels of t that contain only the g-th group or that are deprived of it, respectively. Let

$$\delta_1 = \inf_{1\leqslant g\leqslant|t|} \left\|(I_n - \Pi_{(t^{-g})})\,\mathbb{X}_{(t^{+g})}\,\Delta_{(t^{+g})}^t\,\Omega_y^{-\frac{1}{2}}\right\|_F^2$$

and, for some $K > 0$,

$$\delta_2^K = \inf_{k\,\in\,E_K} \left\|(I_n - \Pi_{(k)})\,\mathbb{X}_{(t)}\,\Delta_{(t)}^t\,\Omega_y^{-\frac{1}{2}}\right\|_F^2$$

with $E_K = \{k \text{ such that } t \not\subset k \text{ and } r_k \leqslant K r_t\}$. Let also,

$$\mu_{n,\,\min}^K = \inf_{k\,\in\,F_K}\; \mu^+\!\left(\frac{\mathbb{X}_{(k)}^t\,\mathbb{X}_{(k)}}{n}\right) \quad \text{and} \quad \bar{\mu}_n = \inf_{k\,\in\,F}\; \mu^*\!\left(\frac{\mathbb{X}_{(k)}^t\,(I_n - \Pi_{(k\cap t)})\,\mathbb{X}_{(k)}}{n}\right)$$

with $F_K = \{k \text{ such that } t \subset k \text{ and } r_k \leqslant (K+1)\, r_t\}$ and $F = \{k \text{ such that } |k \backslash t| > 0\}$, and where, for a square matrix A, $\mu^+(A)$ is the minimum non-zero eigenvalue of A and $\mu^*(A)$ is the geometric mean of the non-zero eigenvalues of A. The hypotheses are those of [31] that we have to slightly adapt. By $f_n \asymp g_n$ we mean that there is a constant $c \neq 0$ such that $f_n/g_n \to c$ as n tends to infinity.

(H.1) There exists a rate such that $m_n = \mathrm{e}^{v_n}$ with $v_n \to +\infty$ and $v_n = o(n)$.
(H.2) The prior slab probability satisfies $1 - \pi_n \asymp 1/m_n$.
(H.3) The shrinkage factors satisfy $n\lambda_n^\sharp \asymp m_n^{2+\eta}\, \bar{\mu}_n^{-\eta}$ and $\mu^K_{n,\min}\, n\lambda_n^\sharp \to +\infty$ for some $\eta > 0$, where $\lambda_n^\sharp = \max_i \lambda_{n,i}$.
(H.4) There exists $\epsilon_1 > 0$ such that $\delta_1 > (1+\epsilon_1)\, r_t\, [(4+\eta) \ln m_n - \eta \ln \bar{\mu}_n]$.
(H.5) There exists $\epsilon_2 > 0$ such that $\delta_2^K > (1+\epsilon_2)\, r_t\, [(4+\eta) \ln m_n - \eta \ln \bar{\mu}_n]$ for some $K > \max(8/\eta + 1, \eta/(\eta - 1))$.

We refer the reader to p. 917 of [31] where the authors give very clarifying comments on the interpretation to be given to these technical assumptions. In particular, while (H.1), (H.2) and (H.3) control the behavior of m_n, π_n and λ_n as n tends to infinity, (H.4) and (H.5) are related to sensitivity and specificity and are therefore in connection with the true model t.

Proposition 3.2. *Suppose that (H.1)–(H.5) are satisfied. Then, as n tends to infinity,*

$$\mathbb{P}(\mathcal{T} \,|\, \mathbb{Y}, \mathbb{X}, \Omega_y) \xrightarrow{\mathbb{P}} 1$$

where $\mathcal{T} = \{t \text{ is selected}\}$ and t is the true model.

Proof. The result is obtained by following the same lines as the proof of Thm 2.1 of [31]. One just has to clarify a few points to solve the issues arising from $q \geqslant 1$ and from the adaptative shrinkage, which is done in Section 5.4. □

Remark 3.1. Obviously, Proposition 3.2 also holds for the sparse setting (with $m = p$) and in that case, it is instructive to draw the parallel with Thm. 1 of [25] even if the estimation procedure is very different. The authors show that, to obtain a $\sqrt{n}$-consistent estimation of the precision matrix Ω in a GGM, Ω must contain at most $\asymp \sqrt{n}/\ln p$ non-zero columns. In the Gibbs sampler (see Proposition 2.1), the slab probability $1 - \pi$ is generated according to a distribution that satisfies

$$\mathbb{E}[1 - \pi \,|\, \Theta_\pi] = \frac{p - N_0 + b}{p + a + b} \quad \text{and} \quad \mathbb{V}(1 - \pi \,|\, \Theta_\pi) = \frac{(N_0 + a)(p - N_0 + b)}{(p + a + b)^2\, (p + a + b + 1)}.$$

Thus, if the model selects $\asymp \sqrt{n}/\ln p$ predictors, it follows that the posterior expectation of $1 - \pi$ is $\asymp \sqrt{n}/(p \ln p) = 1/p$ when $p = \mathrm{e}^{\sqrt{n}}$. In that case, the posterior variance of $1 - \pi$ is $\asymp 1/p^2$. To sum up, in a model with $\asymp \sqrt{n}/\ln p$ predictors selected, the posterior distribution of $1 - \pi$ is very concentrated around $1/p$ which conforms to (H.1) and (H.2). This is not directly comparable due to the different procedures, but it seems interesting to observe that the same orders of magnitude are involved to reach theoretical guarantees for the estimation of Δ.

In the next section, an approach is suggested to deal with sparse-group sparsity in Δ, for a bi-level selection.

4. The sparse-group-sparse setting

To produce a sparse model both at the variable level (for variable selection) and at the group level (for group selection), it seems natural to carry on with our strategy by introducing another spike-and-slab effect into the first one. The predictors are still ordered in m groups of sizes $\kappa_1 + \ldots + \kappa_m = p$. For the g-th group ($1 \leqslant g \leqslant m$), $\lambda_g \in \mathbb{R}$ is the g-th component of $\lambda \in \mathbb{R}^m$ and, for the i-th predictor of this group ($1 \leqslant i \leqslant \kappa_g$), $\nu_{gi} \in \mathbb{R}$ is the i-th component of $\nu_g \in \mathbb{R}^{\kappa_g}$. The i-th column of the covariate submatrix $\mathbb{X}_g$ is $\mathbb{X}_{gi} \in \mathbb{R}^n$ and the corresponding slice of $\underline{\Delta}_g$ is $\Delta_{gi} \in \mathbb{R}^q$ while $\underline{\Delta}_{g\setminus i} \in \mathbb{R}^{q\times(\kappa_g-1)}$ is $\underline{\Delta}_g$ deprived of Δ_{gi}. Here our approach diverges from [29] and [17]. The bi-level selection of the authors is made through spike-and-slab effects both at the group scale and on the individual variances, considered as truncated Gaussians, generating zero groups and (almost surely) zero coefficients within the groups. Let us suggest instead the Bayesian hierarchical model given by

$$
\left\{
\begin{array}{lcl}
\mathbb{Y} \,|\, \mathbb{X}, \Delta, \Omega_y & \sim & \mathcal{MN}_{n\times q}(-\mathbb{X}\,\Delta^t\,\Omega_y^{-1},\, I_n,\, \Omega_y^{-1}) \\
\underline{\Delta}_g \,|\, \nu_g, \lambda_g, \pi & \overset{\perp\!\!\!\perp}{\sim} & (1-\pi_1)\left[(1-\pi_2)\,\mathcal{N}_q(0,\, \lambda_g\,\nu_{gi}\,\Omega_y) + \pi_2\,\delta_0\right]^{\otimes\kappa_g} + \pi_1\,\delta_0 \\
\nu_{gi} & \overset{\perp\!\!\!\perp}{\sim} & \Gamma(\alpha,\, \ell_{gi}) \\
\lambda_g & \overset{\perp\!\!\!\perp}{\sim} & \Gamma(\alpha_g,\, \gamma_g) \\
\Omega_y & \sim & \mathcal{W}_q(u,\, V) \\
\pi_j & \overset{\perp\!\!\!\perp}{\sim} & \beta(a_j,\, b_j)
\end{array}
\right. \tag{4.1}
$$

for $g \in [\![1, m]\!]$, $i \in [\![1, \kappa_g]\!]$ and $j \in [\![1, 2]\!]$, with hyperparameters $\alpha = \frac{1}{2}(q+1)$, $\alpha_g = \frac{1}{2}(q\,\kappa_g+1)$, $\ell_{gi} > 0$, $\gamma_g > 0$, $u > q-1$, $V \in \mathbb{S}^q_{++}$, $a_j > 0$, and $b_j > 0$. In this mixture model, π_1 is the prior spike probability on the groups whereas π_2 is the prior spike probability within the non-zero groups, for a bi-level selection. In terms of cumulative shrinkage effects, λ is an adaptative shrinkage factor acting at the group scale and ν is an adaptative shrinkage factor acting at the predictor scale (λ_g is associated with the direct links between the predictors of group g and all the responses whereas ν_{gi} is associated with the direct links between predictor i of group g and all the responses). In this way, (4.1) opens up many perspectives for dealing with bi-level shrinkage. We can set $\gamma_g = \gamma$ for all g, for a global shrinkage at the group scale. At the predictor scale, when $\ell_{gi} = \ell_g$ for all i, this is a global shrinkage in the g-th group but we might even consider a full global shrinkage $\ell_{gi} = \ell$. However, an identifiability issue may result from the product $\lambda_g\,\nu_{gi}$ between group and within-group effects. Even if the posterior distributions depend on different levels of data that shall resolve it, one can for example fix $\lambda_g = 1$ (for adaptative) or $\nu_{gi} = 1$ (for global) and let the shrinkage entirely rely on the other parameter. Although it achieves the same objectives as those of [29] and [17], this hierarchy seems more consistent with our previous sections (take $\pi_2 = 0$ and $\nu_{gi} = 1$ to remove the within-group effect and recover the group-sparse setting of Section 3, take $\pi_1 = 0$ and $\lambda_g = 1$ to remove the group effect and recover the sparse setting of Section 2). In this context, the degree of sparsity is still characterized by N_0 given in (2.2) for the predictor scale, by G_0 given in (3.2) for the group scale, but also, for the within-group scale, by the number N_{0g} of zero columns in each particular group g, that is, for all $1 \leqslant g \leqslant m$,

$$
N_{0g} = \mathrm{Card}(i,\, \Delta_{gi} = 0) = \sum_{i=1}^{\kappa_g} \mathbb{1}_{\{\Delta_{gi}\,=\,0\}}. \tag{4.2}
$$

We also need to define the number J_0 of zero columns in the non-zero groups, that is

$$J_0 = \mathrm{Card}(i,\, \Delta_{gi} = 0 \text{ and } \underline{\Delta}_g \neq 0) = \sum_{g=1}^{m} N_{0g}\, \mathbb{1}_{\{\Delta_g \neq 0\}}. \tag{4.3}$$

To implement a Gibbs sampler from the full posterior distribution stemming from (4.1), we may use the conditional distributions given in the proposition below.

Proposition 4.1. *In the hierarchical model* (4.1)*, the conditional posterior distributions are as follows.*

– *The parameter* Δ_{gi} *satisfies, for* $g \in [\![1, m]\!]$ *and* $i \in [\![1, \kappa_g]\!]$,

$$\Delta_{gi} \,|\, \Theta_{\Delta_{gi}} \;\sim\; (1 - p_{gi})\, \mathcal{N}_q(-s_{gi}\, H_{gi},\, s_{gi}\, \Omega_y) + p_{gi}\, \delta_0$$

where

$$H_{gi} = \Omega_y\, \mathbb{Y}^t\, \mathbb{X}_{gi} + \sum_{h,j \neq g,i} \langle \mathbb{X}_{gi},\, \mathbb{X}_{hj} \rangle\, \Delta_{hj}, \qquad s_{gi} = \frac{\nu_{gi}\, \lambda_g}{1 + \nu_{gi}\, \lambda_g\, \|\mathbb{X}_{gi}\|^2}$$

and

$$p_{gi} = \frac{\rho_{gi}}{\rho_{gi} + (1 - \pi_1)\,(1 - \pi_2)\,(1 + \nu_{gi}\, \lambda_g\, \|\mathbb{X}_{gi}\|^2)^{-\frac{q}{2}}\, \exp\!\Big(\frac{s_{gi}\, H_{gi}^t\, \Omega_y^{-1}\, H_{gi}}{2}\Big)}$$

in which $\rho_{gi} = (1 - \pi_1)\, \pi_2\, \mathbb{1}_{\{\Delta_{g\setminus i} \neq 0\}} + \pi_1\, \mathbb{1}_{\{\Delta_{g\setminus i} = 0\}}$.

– *The parameter* Ω_y *satisfies*

$$\Omega_y \,|\, \Theta_{\Omega_y} \;\sim\; \mathcal{MGIG}_q\Big(\frac{n - p + N_0 + u}{2},\, \Delta\,(\mathbb{X}^t\, \mathbb{X} + D_{\lambda\nu}^{-1})\, \Delta^t,\, \mathbb{Y}^t\, \mathbb{Y} + V^{-1}\Big)$$

where $D_{\lambda\nu} = \mathrm{diag}(\nu_{11}\, \lambda_1, \ldots, \nu_{1\kappa_1}\, \lambda_1, \ldots, \nu_{m1}\, \lambda_m, \ldots, \nu_{m\kappa_m}\, \lambda_m)$.

– *The parameter* ν *satisfies, for* $g \in [\![1, m]\!]$ *and* $i \in [\![1, \kappa_g]\!]$,

$$\nu_{gi} \,|\, \Theta_{\nu_{gi}} \;\sim\; \mathbb{1}_{\{\Delta_{gi} \neq 0\}}\, \mathcal{GIG}\Big(\frac{1}{2},\, \frac{\Delta_{gi}^t\, \Omega_y^{-1} \Delta_{gi}}{\lambda_g},\, 2\, \ell_{gi}\Big) + \mathbb{1}_{\{\Delta_{gi} = 0\}}\, \Gamma(\alpha,\, \ell_{gi}).$$

– *The parameter* λ *satisfies, for* $g \in [\![1, m]\!]$,

$$\lambda_g \,|\, \Theta_{\lambda_g} \;\sim\; \mathbb{1}_{\{\Delta_g \neq 0\}}\, \mathcal{GIG}\Big(\frac{qN_{0g} + 1}{2},\, \mathrm{tr}(D_{\nu_g}^{-1} \underline{\Delta}_g^t\, \Omega_y^{-1} \underline{\Delta}_g),\, 2\, \gamma_g\Big) + \mathbb{1}_{\{\Delta_g = 0\}}\, \Gamma(\alpha_g,\, \gamma_g)$$

where $D_{\nu_g} = \mathrm{diag}(\nu_{g1}, \ldots, \nu_{g\kappa_g})$.

– *The parameter* π *satisfies, for* $j \in [\![1, 2]\!]$,

$$\pi_j \,|\, \Theta_{\pi_j} \;\sim\; \beta\big(A_j + a_j,\, B_j + b_j\big).$$

where $A_1 = G_0$, $B_1 = m - G_0$, $A_2 = J_0$ *and* $B_2 = p - N_0$.

Proof. See Section 5.3. □

It only remains to give the explicit results for the particular case $q = 1$. The direct links form a row vector such that $\Delta^t \in \mathbb{R}^p$ with groups $\underline{\Delta}_g^t \in \mathbb{R}^{\kappa_g}$ $(1 \leqslant g \leqslant m)$ containing predictors $\Delta_{gi} \in \mathbb{R}$ $(1 \leqslant i \leqslant \kappa_g)$, and the precision matrix of the responses reduces to $\omega_y > 0$. According to the parametrization of the distributions (see Section 1), the corresponding prior distribution of ω_y is $\Gamma(\frac{u}{2}, \frac{1}{2v})$ for $u, v > 0$, like in the other settings, and the one of ν_{gi} is $\mathcal{E}(\ell_{gi})$ for $\ell_{gi} > 0$. The other priors are unchanged.

Corollary 4.1. *In the hierarchical model* (4.1) *with* $q = 1$*, the conditional posterior distributions are as follows.*

– *The parameter* Δ_{gi} *satisfies, for* $g \in [\![1, m]\!]$ *and* $i \in [\![1, \kappa_g]\!]$*,*

$$\Delta_{gi} \,|\, \Theta_{\Delta_{gi}} \;\sim\; (1 - p_{gi})\, \mathcal{N}(-s_{gi}\, h_{gi},\, s_{gi}\, \omega_y) + p_{gi}\, \delta_0$$

where

$$h_{gi} = \omega_y \, \langle \mathbb{X}_{gi},\, \mathbb{Y} \rangle + \sum_{h,j \neq g,i} \langle \mathbb{X}_{gi},\, \mathbb{X}_{hj} \rangle \, \Delta_{hj}, \qquad s_{gi} = \frac{\nu_{gi}\, \lambda_g}{1 + \nu_{gi}\, \lambda_g \, \|\mathbb{X}_{gi}\|^2}$$

and

$$p_{gi} = \frac{\rho_{gi}}{\rho_{gi} + (1 - \pi_1)\,(1 - \pi_2)\,(1 + \nu_{gi}\, \lambda_g \, \|\mathbb{X}_{gi}\|^2)^{-\frac{1}{2}} \, \exp\!\Big(\frac{s_{gi}\, h_{gi}^2}{2\, \omega_y}\Big)}$$

in which $\rho_{gi} = (1 - \pi_1)\, \pi_2 \, \mathbb{1}_{\{\Delta_{g \setminus i} \neq 0\}} + \pi_1 \, \mathbb{1}_{\{\Delta_{g \setminus i} = 0\}}$.

– *The parameter* ω_y *satisfies*

$$\omega_y \,|\, \Theta_{\omega_y} \;\sim\; \mathcal{GIG}\bigg(\frac{n - p + N_0 + u}{2},\, \Delta\,(\mathbb{X}^t\, \mathbb{X} + D_{\lambda\nu}^{-1})\, \Delta^t,\, \mathbb{Y}^t\, \mathbb{Y} + \frac{1}{v}\bigg)$$

where $D_{\lambda\nu} = \mathrm{diag}(\nu_{11}\, \lambda_1, \ldots, \nu_{1\kappa_1}\, \lambda_1, \ldots, \nu_{m1}\, \lambda_m, \ldots, \nu_{m\kappa_m}\, \lambda_m)$.

– *The parameter* ν *satisfies, for* $g \in [\![1, m]\!]$ *and* $i \in [\![1, \kappa_g]\!]$*,*

$$\nu_{gi} \,|\, \Theta_{\nu_{gi}} \;\sim\; \mathbb{1}_{\{\Delta_{gi} \neq 0\}}\, \mathcal{GIG}\bigg(\frac{1}{2},\, \frac{\Delta_{gi}^2}{\lambda_g\, \omega_y},\, 2\, \ell_{gi}\bigg) + \mathbb{1}_{\{\Delta_{gi} = 0\}}\, \mathcal{E}(\ell_{gi}).$$

– *The parameter* λ *satisfies, for* $g \in [\![1, m]\!]$*,*

$$\lambda_g \,|\, \Theta_{\lambda_g} \;\sim\; \mathbb{1}_{\{\Delta_g \neq 0\}}\, \mathcal{GIG}\bigg(\frac{N_{0g} + 1}{2},\, \frac{\underline{\Delta}_g\, D_{\nu_g}^{-1}\, \underline{\Delta}_g^t}{\omega_y},\, 2\, \gamma_g\bigg) + \mathbb{1}_{\{\Delta_g = 0\}}\, \Gamma(\alpha_g,\, \gamma_g)$$

where $D_{\nu_g} = \mathrm{diag}(\nu_{g1}, \ldots, \nu_{g\kappa_g})$.

– *The parameter* π *satisfies, for* $j \in [\![1, 2]\!]$*,*

$$\pi_j \,|\, \Theta_{\pi_j} \;\sim\; \beta\big(A_j + a_j,\, B_j + b_j\big).$$

where $A_1 = G_0$*,* $B_1 = m - G_0$*,* $A_2 = J_0$ *and* $B_2 = p - N_0$*.*

Proof. This is a consequence of Proposition 4.1. □

In the simulation study of Section 6.1, Scen. 5 and 6 are dedicated to the sparse-group-sparse setting. Now, let us prove our assertions by a few computational steps.

5. Conditional posterior distributions

5.1. **The sparse setting: proof of Proposition 2.1.** First of all, the full posterior distribution of the parameters conditional on $\mathbb{X}$ and $\mathbb{Y}$ satisfies

$$\begin{aligned} p(\Delta, \Omega_y, \lambda, \pi \,|\, \mathbb{Y}, \mathbb{X}) \;&\propto\; p(\mathbb{Y} \,|\, \mathbb{X}, \Delta, \Omega_y)\, p(\Delta \,|\, \Omega_y, \lambda, \pi)\, p(\lambda)\, p(\Omega_y)\, p(\pi) \\ &\propto\; |\Omega_y|^{\frac{n}{2}} \, \exp\!\bigg(-\frac{1}{2} \, \Big\| (\mathbb{Y} + \mathbb{X}\, \Delta^t\, \Omega_y^{-1})\, \Omega_y^{\frac{1}{2}} \Big\|_F^2\bigg) \\ &\qquad \times \prod_{i=1}^{p} \bigg[\frac{1 - \pi}{\sqrt{\lambda_i^q \, |\Omega_y|}} \, \exp\!\bigg(-\frac{\Delta_i^t\, \Omega_y^{-1} \Delta_i}{2\, \lambda_i}\bigg) \mathbb{1}_{\{\Delta_i \neq 0\}} \end{aligned}$$

$$+ \pi \, \mathbb{1}_{\{\Delta_i = 0\}} \Bigg] \lambda_i^{\frac{1}{2}(q+1)-1} \, \mathrm{e}^{-\ell_i \, \lambda_i}$$

$$\times \, |\Omega_y|^{\frac{u-q-1}{2}} \, \exp\left(-\frac{\mathrm{tr}(V^{-1}\,\Omega_y)}{2}\right) \pi^{\,a-1} \, (1-\pi)^{\,b-1}. \tag{5.1}$$

On the one hand, exploiting the cyclic property of the trace, a tedious calculation shows that, for all $1 \leqslant i \leqslant p$,

$$\begin{aligned} \left\| (\mathbb{Y} + \mathbb{X}\,\Delta^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}} \right\|_F^2 &= \mathrm{tr}(\mathbb{Y}^t\,\mathbb{Y}\,\Omega_y) + 2\,\mathrm{tr}(\mathbb{X}^t\,\mathbb{Y}\Delta) + \mathrm{tr}(\mathbb{X}^t\,\mathbb{X}\,\Delta^t\,\Omega_y^{-1}\Delta) \\ &= \|\mathbb{X}_i\|^{\,2}\,\Delta_i^t\,\Omega_y^{-1}\Delta_i + 2\sum_{j\neq i}\langle \mathbb{X}_i,\, \mathbb{X}_j\rangle\,\Delta_j^t\,\Omega_y^{-1}\Delta_i + 2\,\mathbb{X}_i^t\,\mathbb{Y}\Delta_i + T_{\neq i} \end{aligned} \tag{5.2}$$

where the term $T_{\neq i}$ does not depend on Δ_i. Thus,

$$\begin{aligned} p(\Delta_i \,|\, \Theta_{\Delta_i}) \;\propto\; & \exp\left(-\frac{1}{2}\,\|\mathbb{X}_i\|^{\,2}\,\Delta_i^t\,\Omega_y^{-1}\Delta_i - \sum_{j\neq i}\langle \mathbb{X}_i,\, \mathbb{X}_j\rangle\,\Delta_j^t\,\Omega_y^{-1}\Delta_i - \mathbb{X}_i^t\,\mathbb{Y}\Delta_i\right) \\ & \times \left[\frac{1-\pi}{\sqrt{\lambda_i^q\,|\Omega_y|}}\,\exp\left(-\frac{\Delta_i^t\,\Omega_y^{-1}\Delta_i}{2\,\lambda_i}\right)\mathbb{1}_{\{\Delta_i\neq 0\}} + \pi\,\mathbb{1}_{\{\Delta_i = 0\}}\right] \\ = \;& \exp\left(-\frac{1}{2}\,(\Delta_i + s_i\,H_i)^{\,t}\,(s_i\,\Omega_y)^{-1}\,(\Delta_i + s_i\,H_i)\right) \\ & \times \exp\left(\frac{s_i\,H_i^t\,\Omega_y^{-1}\,H_i}{2}\right)\frac{1-\pi}{\sqrt{\lambda_i^q\,|\Omega_y|}}\,\mathbb{1}_{\{\Delta_i\neq 0\}} + \pi\,\mathbb{1}_{\{\Delta_i = 0\}} \end{aligned} \tag{5.3}$$

for all $1 \leqslant i \leqslant p$, where

$$H_i = \Omega_y\,\mathbb{Y}^t\,\mathbb{X}_i + \sum_{j\neq i}\langle \mathbb{X}_i,\, \mathbb{X}_j\rangle\,\Delta_j \quad \text{and} \quad s_i = \frac{\lambda_i}{1+\lambda_i\,\|\mathbb{X}_i\|^{\,2}}.$$

This is still a multivariate Gaussian spike-and-slab distribution such that, by renormalizing, the spike has probability

$$p_i = \mathbb{P}(\Delta_i = 0 \,|\, \Theta_{\Delta_i}) = \frac{\pi}{\pi + (1-\pi)\,(1+\lambda_i\,\|\mathbb{X}_i\|^{\,2})^{-\frac{q}{2}}\,\exp\left(\frac{s_i\,H_i^t\,\Omega_y^{-1}H_i}{2}\right)}.$$

On the other hand, coming back to (5.2), we can also write

$$\left\| (\mathbb{Y} + \mathbb{X}\,\Delta^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}} \right\|_F^2 = \mathrm{tr}(\mathbb{Y}^t\,\mathbb{Y}\,\Omega_y) + \mathrm{tr}(\Delta\,\mathbb{X}^t\,\mathbb{X}\,\Delta^t\,\Omega_y^{-1}) + T_{\neq y}$$

where $T_{\neq y}$ does not depend on Ω_y. That leads, *via* (5.1), to

$$\begin{aligned} p(\Omega_y \,|\, \Theta_{\Omega_y}) \;\propto\; & |\Omega_y|^{\frac{n-p+N_0+u-q-1}{2}}\,\exp\Bigg(-\frac{1}{2}\,\mathrm{tr}((\mathbb{Y}^t\,\mathbb{Y} + V^{-1})\,\Omega_y) \\ & - \frac{1}{2}\left(\mathrm{tr}(\Delta\,\mathbb{X}^t\,\mathbb{X}\,\Delta^t\,\Omega_y^{-1}) + \sum_{\Delta_i\neq 0}\frac{\Delta_i^t\,\Omega_y^{-1}\Delta_i}{\lambda_i}\right)\Bigg) \end{aligned}$$

$$= \quad |\Omega_y|^{\frac{n-p+N_0+u-q-1}{2}} \, \exp\!\left(-\frac{1}{2}\,\mathrm{tr}\big((\mathbb{Y}^t\,\mathbb{Y}+V^{-1})\,\Omega_y+\Delta\,(\mathbb{X}^t\,\mathbb{X}+D_\lambda^{-1})\,\Delta^t\,\Omega_y^{-1}\big)\right) \tag{5.4}$$

where N_0 is given in (2.2) and $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. Finally, it is easy to see that, for all $1 \leqslant i \leqslant p$,

$$p(\lambda_i \,|\, \Theta_{\lambda_i}) \;\propto\; \frac{1}{\sqrt{\lambda_i}}\,\exp\!\left(-\frac{\underline{\Delta}_i^t\,\Omega_y^{-1}\underline{\Delta}_i}{2\,\lambda_i}-\ell_i\,\lambda_i\right)\mathbb{1}_{\{\Delta_i\neq 0\}} + \lambda_i^{\frac{1}{2}(q+1)-1}\,\mathrm{e}^{-\ell_i\,\lambda_i}\,\mathbb{1}_{\{\Delta_i=0\}} \tag{5.5}$$

whereas

$$p(\pi \,|\, \Theta_\pi) \;\propto\; \pi^{N_0+a-1}\,(1-\pi)^{p-N_0+b-1}. \tag{5.6}$$

We recognize in (5.3), (5.4), (5.5) and (5.6) the announced conditional posterior distributions, which concludes the proof. □

5.2. **The group-sparse setting: proof of Proposition 3.1.** The full posterior distribution of the parameters conditional on $\mathbb{X}$ and $\mathbb{Y}$ satisfies

$$\begin{aligned}
p(\Delta, \Omega_y, \lambda, \pi \,|\, \mathbb{Y}, \mathbb{X}) \;&\propto\; p(\mathbb{Y} \,|\, \mathbb{X}, \Delta, \Omega_y)\, p(\Delta \,|\, \Omega_y, \lambda, \pi)\, p(\lambda)\, p(\Omega_y)\, p(\pi) \\
&\propto\; |\Omega_y|^{\frac{n}{2}}\,\exp\!\left(-\frac{1}{2}\,\left\|(\mathbb{Y}+\mathbb{X}\,\Delta^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}}\right\|_F^2\right) \\
&\quad\times \prod_{g=1}^{m}\left[\frac{1-\pi}{\sqrt{\lambda_g^{q\,\kappa_g}\,|\Omega_y|^{\kappa_g}}}\,\exp\!\left(-\frac{\mathrm{tr}(\underline{\Delta}_g^t\,\Omega_y^{-1}\underline{\Delta}_g)}{2\,\lambda_g}\right)\mathbb{1}_{\{\Delta_g\neq 0\}}\right. \\
&\qquad\left. +\;\pi\,\mathbb{1}_{\{\Delta_g=0\}}\right]\lambda_g^{\frac{1}{2}(q\,\kappa_g+1)-1}\,\mathrm{e}^{-\ell_g\,\lambda_g} \\
&\quad\times |\Omega_y|^{\frac{u-q-1}{2}}\,\exp\!\left(-\frac{\mathrm{tr}(V^{-1}\,\Omega_y)}{2}\right)\pi^{\,a-1}\,(1-\pi)^{\,b-1}.
\end{aligned} \tag{5.7}$$

Like in the previous proof, a first important step is to note that, for all $1 \leqslant g \leqslant m$,

$$\begin{aligned}
\left\|(\mathbb{Y}+\mathbb{X}\,\Delta^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}}\right\|_F^2 \;&=\; \left\|\mathbb{Y}\,\Omega_y^{\frac{1}{2}}+\sum_{j=1}^{m}\mathbb{X}_j\,\underline{\Delta}_j^t\,\Omega_y^{-\frac{1}{2}}\right\|_F^2 \\
&=\; \|\mathbb{X}_g\,\underline{\Delta}_g^t\,\Omega_y^{-\frac{1}{2}}\|_F^2 + 2\sum_{j\neq g}\mathrm{tr}(\underline{\Delta}_j\,\mathbb{X}_j^t\,\mathbb{X}_g\,\underline{\Delta}_g^t\,\Omega_y^{-1}) \\
&\qquad +\;2\,\mathrm{tr}(\mathbb{X}_g^t\,\mathbb{Y}\,\underline{\Delta}_g)+T_{\neq g}
\end{aligned} \tag{5.8}$$

where the term $T_{\neq g}$ does not depend on $\underline{\Delta}_g$. Thus, after a tedious calculation exploiting the cyclic property of the trace, one can obtain the factorization

$$\begin{aligned}
p(\underline{\Delta}_g \,|\, \Theta_{\Delta_g}) \;&\propto\; \exp\!\left(-\frac{1}{2}\,\|\mathbb{X}_g\,\underline{\Delta}_g^t\,\Omega_y^{-\frac{1}{2}}\|_F^2-\sum_{j\neq g}\mathrm{tr}(\underline{\Delta}_j\,\mathbb{X}_j^t\,\mathbb{X}_g\,\underline{\Delta}_g^t\,\Omega_y^{-1})-\mathrm{tr}(\mathbb{X}_g^t\,\mathbb{Y}\,\underline{\Delta}_g)\right) \\
&\quad\times\left[\frac{1-\pi}{\sqrt{\lambda_g^{q\,\kappa_g}\,|\Omega_y|^{\kappa_g}}}\,\exp\!\left(-\frac{\mathrm{tr}(\underline{\Delta}_g^t\,\Omega_y^{-1}\underline{\Delta}_g)}{2\,\lambda_g}\right)\mathbb{1}_{\{\Delta_g\neq 0\}}+\pi\,\mathbb{1}_{\{\Delta_g=0\}}\right] \\
&=\; \exp\!\left(-\frac{1}{2}\,\mathrm{tr}\big(S_g^{-1}(\underline{\Delta}_g+H_g\,S_g)^t\,\Omega_y^{-1}(\underline{\Delta}_g+H_g\,S_g)\big)\right)
\end{aligned}$$

$$\times\ \exp\left(\frac{\mathrm{tr}(H_g^t\,\Omega_y^{-1}\,H_g\,S_g)}{2}\right)\frac{1-\pi}{\sqrt{\lambda_g^{q\,\kappa_g}\,|\Omega_y|^{\kappa_g}}}\,\mathbb{1}_{\{\Delta_g\neq 0\}}+\pi\,\mathbb{1}_{\{\Delta_g=0\}} \tag{5.9}$$

for all $1\leqslant g\leqslant m$, where

$$H_g=\Omega_y\,\mathbb{Y}^t\,\underline{\mathbb{X}}_g+\sum_{j\neq g}\underline{\Delta}_j\,\underline{\mathbb{X}}_j^t\,\underline{\mathbb{X}}_g\quad\text{ and }\quad S_g=\lambda_g\big(I_{\kappa_g}+\lambda_g\,\underline{\mathbb{X}}_g^t\,\underline{\mathbb{X}}_g\big)^{-1}.$$

We recognize the announced Gaussian spike-and-slab distribution, and the probability of the spike is given, after renormalization, by

$$p_g=\mathbb{P}(\underline{\Delta}_g=0\,|\,\Theta_{\Delta_g})=\frac{\pi}{\pi+(1-\pi)\,|I_{\kappa_g}+\lambda_g\,\underline{\mathbb{X}}_g^t\,\underline{\mathbb{X}}_g|^{-\frac{q}{2}}\,\exp\Big(\frac{\mathrm{tr}(H_g^t\,\Omega_y^{-1}H_g\,S_g)}{2}\Big)}.$$

Following the same lines as the ones used to establish (5.4), we obtain from (5.7) the conditional distribution

$$\begin{aligned}p(\Omega_y\,|\,\Theta_{\Omega_y})\ &\propto\ |\Omega_y|^{\frac{n-p+N_0+u-q-1}{2}}\,\exp\bigg(-\frac{1}{2}\,\mathrm{tr}((\mathbb{Y}^t\,\mathbb{Y}+V^{-1})\,\Omega_y)\\&\qquad-\frac{1}{2}\bigg(\mathrm{tr}(\Delta\,\mathbb{X}^t\,\mathbb{X}\,\Delta^t\,\Omega_y^{-1})+\sum_{\Delta_g\neq 0}\frac{\mathrm{tr}(\underline{\Delta}_g^t\,\Omega_y^{-1}\underline{\Delta}_g)}{\lambda_g}\bigg)\bigg)\\&=\ |\Omega_y|^{\frac{n-p+N_0+u-q-1}{2}}\,\exp\Big(-\frac{1}{2}\,\mathrm{tr}\big((\mathbb{Y}^t\,\mathbb{Y}+V^{-1})\,\Omega_y+\Delta\,(\mathbb{X}^t\,\mathbb{X}+D_\lambda^{-1})\,\Delta^t\,\Omega_y^{-1}\big)\Big)\end{aligned} \tag{5.10}$$

where $D_\lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_1,\ldots,\lambda_m,\ldots,\lambda_m)$ with each λ_g duplicated κ_g times, and since we can note that, due to the continuous nature of $\Delta\,|\{\Delta\neq 0\}$,

$$\sum_{g=1}^m\kappa_g\mathbb{1}_{\{\Delta_g\neq 0\}}=p-N_0$$

for N_0 given in (2.2). Next, we obtain in a simpler way that, for all $1\leqslant g\leqslant m$,

$$\begin{aligned}p(\lambda_g\,|\,\Theta_{\lambda_g})\ &\propto\ \frac{1}{\sqrt{\lambda_g}}\,\exp\bigg(-\frac{\mathrm{tr}(\underline{\Delta}_g^t\,\Omega_y^{-1}\underline{\Delta}_g)}{2\,\lambda_g}-\ell_g\,\lambda_g\bigg)\,\mathbb{1}_{\{\Delta_g\neq 0\}}\\&\qquad+\lambda_g^{\frac{1}{2}(q\,\kappa_g+1)-1}\,\mathrm{e}^{-\ell_g\,\lambda_g}\,\mathbb{1}_{\{\Delta_g=0\}}.\end{aligned} \tag{5.11}$$

Finally,

$$p(\pi\,|\,\Theta_\pi)\ \propto\ \pi^{G_0+a-1}\,(1-\pi)^{m-G_0+b-1} \tag{5.12}$$

where G_0 is defined in (3.2). We can check that the conditional distributions (5.9), (5.10), (5.11) and (5.12) correspond to the ones announced in the proposition, which concludes the proof. □

5.3. **The sparse-group-sparse setting: proof of Proposition 4.1.** The full posterior distribution of the parameters conditional on $\mathbb{X}$ and $\mathbb{Y}$ satisfies

$$\begin{aligned}p(\Delta,\Omega_y,\nu,\lambda,\pi\,|\,\mathbb{Y},\mathbb{X})\ &\propto\ p(\mathbb{Y}\,|\,\mathbb{X},\Delta,\Omega_y)\,p(\Delta\,|\,\Omega_y,\nu,\lambda,\pi)\,p(\nu)\,p(\lambda)\,p(\Omega_y)\,p(\pi)\\&\propto\ |\Omega_y|^{\frac{n}{2}}\,\exp\bigg(-\frac{1}{2}\,\Big\|(\mathbb{Y}+\mathbb{X}\,\Delta^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}}\Big\|_F^2\bigg)\end{aligned}$$

$$\times \prod_{g=1}^{m} \Bigg[\Big((1-\pi_1)\, P_g\, \mathbb{1}_{\{\Delta_g \neq 0\}} + \pi_1\, \mathbb{1}_{\{\Delta_g = 0\}} \Big) \times \lambda_g^{\frac{1}{2}(q\,\kappa_g+1)-1}\, \mathrm{e}^{-\gamma_g\,\lambda_g} \prod_{i=1}^{\kappa_g} \nu_{gi}^{\frac{1}{2}(q+1)-1}\, \mathrm{e}^{-\ell_{gi}\,\nu_{gi}} \Bigg]$$

$$\times\, |\Omega_y|^{\frac{u-q-1}{2}}\, \exp\!\left(-\frac{\mathrm{tr}(V^{-1}\,\Omega_y)}{2}\right) \prod_{j=1}^{2} \pi_j^{\,a_j-1}\,(1-\pi_j)^{\,b_j-1} \tag{5.13}$$

where, for $1 \leqslant g \leqslant m$,

$$P_g = \prod_{i=1}^{\kappa_g} \left[\frac{1-\pi_2}{\sqrt{(\nu_{gi}\,\lambda_g)^{\,q}\,|\Omega_y|}}\, \exp\!\left(-\frac{\Delta_{gi}^t\,\Omega_y^{-1}\Delta_{gi}}{2\,\nu_{gi}\,\lambda_g}\right) \mathbb{1}_{\{\Delta_{gi}\neq 0\}} + \pi_2\, \mathbb{1}_{\{\Delta_{gi}=0\}} \right].$$

Using the same decompositions as (5.2) or (5.8), the full posterior distribution given above leads to

$$\begin{aligned}
p(\Delta_{gi}\,|\,\Theta_{\Delta_{gi}}) \;\propto\; & \exp\!\left(-\frac{1}{2}\,\|\mathbb{X}_{gi}\|^{\,2}\,\Delta_{gi}^t\,\Omega_y^{-1}\Delta_{gi} - \sum_{h,j\neq g,i} \langle \mathbb{X}_{gi},\, \mathbb{X}_{hj}\rangle\,\Delta_{hj}^t\,\Omega_y^{-1}\Delta_{gi} - \mathbb{X}_{gi}^t\,\mathbb{Y}\Delta_{gi}\right) \\
& \times \left[(1-\pi_1) \left[\frac{1-\pi_2}{\sqrt{(\nu_{gi}\,\lambda_g)^{\,q}\,|\Omega_y|}}\, \exp\!\left(-\frac{\Delta_{gi}^t\,\Omega_y^{-1}\Delta_{gi}}{2\,\nu_{gi}\,\lambda_g}\right) \mathbb{1}_{\{\Delta_{gi}\neq 0\}} \right.\right. \\
& \qquad \left.\left. + \,\pi_2\, \mathbb{1}_{\{\Delta_{gi}=0\}} \right] \mathbb{1}_{\{\Delta_g\neq 0\}} + \pi_1\, \mathbb{1}_{\{\Delta_g = 0\}} \right] \\
= \;& \exp\!\left(-\frac{1}{2}\,(\Delta_{gi} + s_{gi}\,H_{gi})^{\,t}\,(s_{gi}\,\Omega_y)^{-1}\,(\Delta_{gi} + s_{gi}\,H_{gi})\right) \\
& \times\, \exp\!\left(\frac{s_{gi}\,H_{gi}^t\,\Omega_y^{-1}\,H_{gi}}{2}\right) \frac{(1-\pi_1)\,(1-\pi_2)}{\sqrt{(\nu_{gi}\,\lambda_g)^{\,q}\,|\Omega_y|}}\, \mathbb{1}_{\{\Delta_{gi}\neq 0\}} \\
& + \Big((1-\pi_1)\,\pi_2\, \mathbb{1}_{\{\Delta_{g\setminus i}\neq 0\}} + \pi_1\, \mathbb{1}_{\{\Delta_{g\setminus i}=0\}} \Big)\, \mathbb{1}_{\{\Delta_{gi}=0\}}
\end{aligned} \tag{5.14}$$

for $1 \leqslant g \leqslant m$ and $1 \leqslant i \leqslant \kappa_g$, where $\Delta_{g\setminus i}$ is Δ_g deprived of Δ_{gi},

$$H_{gi} = \Omega_y\,\mathbb{Y}^t\,\mathbb{X}_{gi} + \sum_{h,j\neq g,i} \langle \mathbb{X}_{gi},\, \mathbb{X}_{hj}\rangle\,\Delta_{hj} \quad \text{and} \quad s_{gi} = \frac{\nu_{gi}\,\lambda_g}{1+\nu_{gi}\,\lambda_g\,\|\mathbb{X}_{gi}\|^{\,2}}.$$

Here, we used the binary equalities stemming from $\{\Delta_{gi}\neq 0\}\cap\{\Delta_g\neq 0\} = \{\Delta_{gi}\neq 0\}$, $\{\Delta_{gi}= 0\}\cap\{\Delta_g\neq 0\} = \{\Delta_{gi}= 0\}\cap\{\Delta_{g\setminus i}\neq 0\}$ and $\{\Delta_{gi}= 0\}\cap\{\Delta_g= 0\} = \{\Delta_{gi}= 0\}\cap\{\Delta_{g\setminus i}= 0\}$, which turn out to be very useful to separate Δ_{gi} and $\Theta_{\Delta_{gi}}$. This is characteristic of a multivariate Gaussian spike-and-slab distribution. By renormalizing, one can see that the spike has probability

$$p_{gi} = \mathbb{P}(\Delta_{gi}=0\,|\,\Theta_{\Delta_{gi}}) = \frac{\rho_{gi}}{\rho_{gi} + (1-\pi_1)\,(1-\pi_2)\,(1+\nu_{gi}\,\lambda_g\,\|\mathbb{X}_{gi}\|^{\,2})^{-\frac{q}{2}}\, \exp\!\left(\frac{s_{gi}\,H_{gi}^t\,\Omega_y^{-1}\,H_{gi}}{2}\right)}$$

with

$$\rho_{gi} = (1-\pi_1)\,\pi_2\, \mathbb{1}_{\{\Delta_{g\setminus i}\neq 0\}} + \pi_1\, \mathbb{1}_{\{\Delta_{g\setminus i}=0\}}.$$

Next, following (5.13) and the reasoning used to establish (5.4), we may also write

$$p(\Omega_y \,|\, \Theta_{\Omega_y}) \;\propto\; |\Omega_y|^{\frac{n-p+N_0+u-q-1}{2}} \exp\biggl(-\frac{1}{2}\operatorname{tr}((\mathbb{Y}^t\,\mathbb{Y} + V^{-1})\,\Omega_y) - \frac{1}{2}\biggl(\operatorname{tr}(\Delta\,\mathbb{X}^t\,\mathbb{X}\,\Delta^t\,\Omega_y^{-1}) + \sum_{\Delta_{gi}\neq 0} \frac{\Delta_{gi}^t\,\Omega_y^{-1}\Delta_{gi}}{\nu_{gi}\,\lambda_g}\biggr)\biggr)$$

$$= |\Omega_y|^{\frac{n-p+N_0+u-q-1}{2}} \exp\Bigl(-\frac{1}{2}\operatorname{tr}\bigl((\mathbb{Y}^t\,\mathbb{Y} + V^{-1})\,\Omega_y + \Delta\,(\mathbb{X}^t\,\mathbb{X} + D_{\lambda\nu}^{-1})\,\Delta^t\,\Omega_y^{-1}\bigr)\Bigr) \tag{5.15}$$

where N_0 is given in (2.2) and $D_{\lambda\nu} = \operatorname{diag}(\nu_{11}\lambda_1, \ldots, \nu_{1\kappa_1}\lambda_1, \ldots, \nu_{m1}\lambda_m, \ldots, \nu_{m\kappa_m}\lambda_m)$. The shrinkage parameters ν and λ are easier to handle. For $1 \leqslant g \leqslant m$ and $1 \leqslant i \leqslant \kappa_g$,

$$p(\nu_{gi} \,|\, \Theta_{\nu_{gi}}) \;\propto\; \frac{1}{\sqrt{\nu_{gi}}} \exp\Bigl(-\frac{\Delta_{gi}^t\,\Omega_y^{-1}\Delta_{gi}}{2\,\nu_{gi}\,\lambda_g} - \ell_{gi}\,\nu_{gi}\Bigr)\mathbb{1}_{\{\Delta_{gi}\neq 0\}} + \nu_{gi}^{\frac{1}{2}(q+1)-1}\,\mathrm{e}^{-\ell_{gi}\,\nu_{gi}}\,\mathbb{1}_{\{\Delta_{gi}=0\}} \tag{5.16}$$

whereas

$$p(\lambda_g \,|\, \Theta_{\lambda_g}) \;\propto\; \lambda_g^{\frac{qN_{0g}-1}{2}} \exp\Bigl(-\frac{\operatorname{tr}(D_{\nu_g}^{-1}\underline{\Delta}_g^t\,\Omega_y^{-1}\underline{\Delta}_g)}{2\,\lambda_g} - \gamma_g\,\lambda_g\Bigr)\mathbb{1}_{\{\Delta_g\neq 0\}} + \lambda_g^{\frac{1}{2}(q\,\kappa_g+1)-1}\,\mathrm{e}^{-\gamma_g\,\lambda_g}\,\mathbb{1}_{\{\Delta_g=0\}} \tag{5.17}$$

where N_{0g} is defined in (4.2) and $D_{\nu_g} = \operatorname{diag}(\nu_{g1}, \ldots, \nu_{g\kappa_g})$. Finally,

$$p(\pi_1 \,|\, \Theta_{\pi_1}) \;\propto\; \pi_1^{G_0+a_1-1}\,(1-\pi_1)^{m-G_0+b_1-1} \tag{5.18}$$

and

$$p(\pi_2 \,|\, \Theta_{\pi_2}) \;\propto\; \pi_2^{J_0+a_2-1}\,(1-\pi_2)^{p-N_0+b_2-1} \tag{5.19}$$

where G_0 and J_0 are given in (3.2) and (4.3), respectively. For the latter result, we used the fact that the number of non-zero columns in the non-zero groups must coincide with the number of non-zero columns of Δ, that is $p - N_0$. Like in the previous proofs, we recognize the announced conditional distributions in (5.14), (5.15), (5.16), (5.17), (5.18) and (5.19). That concludes these tedious calculations. □

5.4. **Proof of Proposition 3.2.** The result is obtained by following the steps of the proof of Thm 2.1 in [31] but, beforehand, we need to clarify a few points to extend the reasoning of the authors from $q = 1$ to $q \geqslant 1$ and take into account the adaptative shrinkage. For any model k, let $\mathcal{K} = \{k \text{ is selected}\}$ so that $\mathcal{K} = \mathcal{T}$ when the true model t is considered. First, recall that λ and π are fixed and rewrite (5.7) like

$$\mathbb{P}_\Delta(\mathcal{K} \,|\, \mathbb{Y}, \mathbb{X}, \Omega_y) \;\propto\; \exp\Bigl(-\frac{1}{2}\,\Bigl\|(\mathbb{Y} + \mathbb{X}_{(k)}\,\Delta_{(k)}^t\,\Omega_y^{-1})\,\Omega_y^{\frac{1}{2}}\Bigr\|_F^2\Bigr) \times \frac{(1-\pi)^{|k|}}{\pi^{|k|}\,\sqrt{|\Lambda_k|^q\,|\Omega_y|^{k_r}}} \exp\Bigl(-\frac{\operatorname{tr}(\Delta_{(k)}^t\,\Omega_y^{-1}\Delta_{(k)}\,D_k^{-1})}{2}\Bigr)$$

$$\propto \frac{(1-\pi)^{|k|}}{\pi^{|k|}\sqrt{|\Lambda_k|^q\,|\Omega_y|^{k_r}}}\,\exp\left(\frac{\mathrm{tr}(\widetilde{\Delta}_{(k)}\,F_k\,\widetilde{\Delta}_{(k)}^t\,\Omega_y^{-1})}{2}\right)$$

$$\times\ \exp\left(-\frac{1}{2}\,\mathrm{tr}\Big((\Delta_{(k)}-\widetilde{\Delta}_{(k)})\,F_k\,(\Delta_{(k)}-\widetilde{\Delta}_{(k)})^t\,\Omega_y^{-1}\Big)\right) \tag{5.20}$$

where $F_k = D_k^{-1} + \mathbb{X}_{(k)}^t\,\mathbb{X}_{(k)}$, $D_k = \mathrm{diag}((\lambda_\ell,\ldots,\lambda_\ell)_{\ell\in k})$ with each λ_ℓ duplicated κ_ℓ times, $k_r = \|(\kappa_\ell)_{\ell\in k}\|_1$, $\Lambda_k = \mathrm{diag}((\lambda_\ell^{\kappa_\ell})_{\ell\in k})$ and

$$\widetilde{\Delta}_{(k)} = -\Omega_y\,\mathbb{Y}^t\,\mathbb{X}_{(k)}\,F_k^{-1}.$$

Then, integrating over $\Delta_{(k)}$, it follows (see Def. 1.1 with $\Sigma_1 = \Omega_y$ and $\Sigma_2 = F_k^{-1}$) that

$$\begin{aligned}
\mathbb{P}(\mathcal{K}\,|\,\mathbb{Y},\mathbb{X},\Omega_y) &= \int_{\mathbb{R}^{q\times\kappa_k}} \mathbb{P}_\Delta(\mathcal{K}\,|\,\mathbb{Y},\mathbb{X},\Omega_y,\lambda,\pi)\,\mathrm{d}\Delta_{(k)}\\
&\propto \frac{(1-\pi)^{|k|}}{\pi^{|k|}\sqrt{|\Lambda_k|^q\,|F_k|^q}}\,\exp\left(\frac{\mathrm{tr}(\widetilde{\Delta}_{(k)}\,F_k\,\widetilde{\Delta}_{(k)}^t\,\Omega_y^{-1})}{2}\right)\\
&\propto \left(\frac{1-\pi}{\pi}\right)^{|k|}|\Lambda_k|^{-\frac{q}{2}}\,|F_k|^{-\frac{q}{2}}\,\exp\left(-\frac{1}{2}\,\mathrm{tr}\big(\mathbb{Y}^{*\,t}\,(I_n-\mathbb{X}_{(k)}\,F_k^{-1}\,\mathbb{X}_{(k)}^t)\,\mathbb{Y}^*\big)\right)\\
&= \left(\frac{1-\pi}{\pi}\right)^{|k|}|\Lambda_k|^{-\frac{q}{2}}\,|F_k|^{-\frac{q}{2}}\,\exp\left(-\frac{1}{2}\,\left(\mathrm{RSS}_k(\widetilde{\Delta}_{(k)}^*)+\big\|\widetilde{\Delta}_{(k)}^*\,D_k^{-\frac{1}{2}}\big\|_F^2\right)\right)
\end{aligned}$$

where $\mathbb{Y}^* = \mathbb{Y}\,\Omega_y^{\frac{1}{2}}$, $\widetilde{\Delta}_{(k)}^* = \Omega_y^{-\frac{1}{2}}\,\widetilde{\Delta}_{(k)}$ and $\mathrm{RSS}_k : H\in\mathbb{R}^{q\times\kappa_k} \mapsto \|\mathbb{Y}^*-\mathbb{X}_{(k)}\,H^{\,t}\|_F^2$ is the residual sum of squares function in the renormalized linear model indexed by k, that is

$$\mathbb{Y}^* = -\mathbb{X}_{(k)}\,\Delta_{(k)}^t\,\Omega_y^{-\frac{1}{2}} + E^*$$

with $E^* = E\,\Omega_y^{\frac{1}{2}} \sim \mathcal{MN}_{n\times q}(0,\,I_n,\,I_q)$. Thus, the so-called posterior ratio between any false model k and t is given by

$$\mathrm{PR}(k,t) = \frac{\mathbb{P}(\mathcal{K}\,|\,\mathbb{Y},\mathbb{X},\Omega_y)}{\mathbb{P}(\mathcal{T}\,|\,\mathbb{Y},\mathbb{X},\Omega_y)} = \frac{Q_k}{Q_t}\left(\frac{1-\pi}{\pi}\right)^{|k|-|t|}\mathrm{e}^{-\frac{1}{2}\,(\widetilde{R}_k-\widetilde{R}_t)}$$

with $Q_k = |\Lambda_k|^{-\frac{q}{2}}\,|F_k|^{-\frac{q}{2}}$ and $\widetilde{R}_k = \mathrm{RSS}_k(\widetilde{\Delta}_{(k)}^*) + \|\widetilde{\Delta}_{(k)}^*\,D_k^{-\frac{1}{2}}\|_F^2$, using the notation of [31]. In particular, due to the generalized ridge penalty,

$$\widetilde{\Delta}_{(k)}^* = \arg\min_H\ \left(\mathrm{RSS}_k(H) + \big\|H D_k^{-\frac{1}{2}}\big\|_F^2\right) \tag{5.21}$$

so that for nested models k_1 and k_2 (with $k_1\subseteq k_2$), we must have $\widetilde{R}_{k_2} \leqslant \widetilde{R}_{k_1}$. Let also $R_k = \|(I_n-\Pi_{(k)})\,\mathbb{Y}^*\|_F^2 = \|(I_q\otimes(I_n-\Pi_{(k)}))\,\mathrm{vec}(\mathbb{Y}^*)\|_2^2$. Cochran's theorem entails the chi-squared distributions $R_t \sim \chi^2(q\,(n-r_t))$ and $R_t - R_k \sim \chi^2(q\,(r_k-r_t))$ for any 'bigger' model $k\supset t$ and $q\geqslant 1$. Combining all these preliminary considerations, the strategy of [31] now applies and leads, under our revised hypotheses, to

$$\frac{1-\mathbb{P}(\mathcal{T}\,|\,\mathbb{Y},\mathbb{X},\Omega_y)}{\mathbb{P}(\mathcal{T}\,|\,\mathbb{Y},\mathbb{X},\Omega_y)} = \sum_{k\neq t}\mathrm{PR}(k,t)\ \xrightarrow{\ \mathbb{P}\ }\ 0.$$

□

6. Empirical results

In this section, let us call (s), (gs) and (sgs) the related settings, and let us denote by (ad) the adaptative shrinkage and by (gl) the global shrinkage. First of all, these models contain many hyperparameters that have to be carefully tuned. Our experiments showed that, unsurprinsingly, the results are strongly impacted by the prior amount of shrinkage on Δ, driven by ℓ and even by γ for (sgs). Apart from the usual cross-validation procedures, we could stay in line with our Bayesian approach and suggest conjugate Gamma hyperpriors. This is very easy to implement, but the hyperparameters are now replaced by other hyperparameters and the same questions arise. Instead, like in [29] and [17], we follow the idea of [22] and we use a Monte-Carlo EM algorithm. By way of example, from the full posterior probability (5.1) and since $\lambda_i \sim \Gamma(\alpha,\, \ell_i)$ for all i, it is not hard to see that, with (s),

$$\ln p(\Delta, \Omega_y, \lambda, \pi \,|\, \mathbb{Y}, \mathbb{X}) = \sum_{i=1}^{p} (\alpha \, \ln \ell_i - \ell_i \, \lambda_i) + T_{\neq \ell}$$

where the term $T_{\neq \ell}$ does not depend on ℓ. Thus, the k-th iteration of the EM algorithm should lead to

$$\ell_i^{(k)} = \frac{\frac{1}{2}(q+1)}{\mathbb{E}^{(k-1)}[\lambda_i \,|\, \mathbb{Y},\, \mathbb{X}]} \quad \text{and} \quad \ell^{(k)} = \frac{\frac{p}{2}(q+1)}{\sum_{i=1}^{p} \mathbb{E}^{(k-1)}[\lambda_i \,|\, \mathbb{Y},\, \mathbb{X}]}$$

for the adaptative shrinkage and the global shrinkage ($\lambda_i = \lambda$), respectively. The intractable conditional expectations are then estimated with the help of the Gibbs samples. For (gs), the results are mainly the same as above (replace $q+1$ by $q\kappa_g+1$ in the first case, $p(q+1)$ by $qp+m$ in the second case and consider $1 \leqslant g \leqslant m$ instead of $1 \leqslant i \leqslant p$), and similar results also follow with (sgs). Recall that our definitions of the adaptative and global shrinkages are given in the corresponding sections, in the description of the hierarchical models. The tuning of u and V (or v) is actually trickier. Because $\mathbb{E}[\Omega_y] = uV$, we set $V = \frac{1}{u} I_q$ and u is conveniently chosen to be the smallest integer such that Ω_y is (almost surely) invertible, that is $u = q$ (see *e.g.* [3]). This is particularly adapted when the dataset is standardized. Finally, a and b reflect the degree of sparsity to introduce in the direct links. We can set $a \gg b$ to promote sparse settings, which is potentially interesting when $p \gg n$, but $a = b = 1$ is a standard non-informative choice and $a < b$ may also be useful for variable selection (see *e.g.* the real dataset of Section 6.2). They can be chosen from a cross-validation step (for prediction purposes) or to enforce some degree of sparsity (for selection purposes), just like a practitioner manages the tuning parameter of the Lasso. The posterior median is used to estimate Δ and get sparsity whereas the posterior mean is used to estimate Ω_y. Indeed, we don't want to impose any sparsity on Ω_y (q is small), so we decided to retain this standard choice. But the concern is much greater for Δ because some coordinates must be exactly zero. This is the reason why the posterior median seemed a more appropriate choice (in particular, it suffices for the sampler to generate zeros more than half the time for the empirical posterior median to be zero). Due to the huge amount of calculations in the simulations, the estimations are made on the basis of 3000 iterations of the sampler in which the first 2000 are burn-ins. This is revised upwards for the real data (10000 iterations with 5000 burn-ins).

Remark 6.1. To the best of our knowledge, there is no simple way to sample from the $\mathcal{MGIG}_d$ distribution as soon as $d > 1$. The recent method described in Sec. 3.3.2 of [8], relying on

the Matsumoto-Yor property (see Thm. 3.1 of [19]) to get a $\mathcal{MGIG}_d$ sample from the very standard $\mathcal{GIG}$ and $\mathcal{W}_d$ distributions, is unfortunately inapplicable in our context. Indeed, for example in the sparse setting, that would require finding $z \in \mathbb{R}^q$ such that $\mathbb{Y}^t\,\mathbb{Y} + V^{-1} = b\,zz^t$ for some $b > 0$, which is clearly impossible since $\mathbb{Y}^t\,\mathbb{Y} + V^{-1}$ has full rank. In [9], the authors show that $\mathcal{MGIG}_d(\nu,\, A,\, B)$ is a unimodal distribution of which mode $M \in \mathbb{S}_{++}^d$ is the unique solution of the algebraic Riccati equation $(d + 1 - 2\,\nu)\,M + MBM = A$, and a standard importance sampling approach follows for the mean of the distribution. Our fallback solution is to solve this Riccati equation at each step and to replace all $\mathcal{MGIG}_d$ random variables by the (unique) mode of the consecutive distributions. To assess the credibility of this *ad hoc* sampling, the 'oracle' models in which Ω_y and the shrinkage parameters are known are added to the simulations. We will see that, despite an unavoidable loss, the results remain pretty consistent. In particular, the support recovery does not appear to be impacted.

6.1. **A simulation study.** In this empirical section[1], the matrix of order $d \geqslant 1$ given by

$$C_d = \left(\rho^{|i-j|}\right)_{1 \leqslant i,j \leqslant d}$$

will be used as a typical covariance structure, for some $0 \leqslant \rho < 1$. Thus, the precision matrices will be chosen as a multiple of C_d^{-1} to keep the same guideline in our simulations. The responses

$$Y_k = B^t\,X_k + E_k$$

are generated through relations (1.1) where, for all $1 \leqslant k \leqslant n$, $E_k \sim \mathcal{N}(0, R)$. Because our models assume prior independence (or group-independence) in the columns of Δ, it seems necessary to look at the influence of correlation among the predictors. So the standard choice $X_k \sim \mathcal{N}(0, I_p)$ is first considered, but in some cases we will also test $X_k \sim \mathcal{N}(0, C_p)$ for $\rho = 0.5$ and $\rho = 0.9$ to introduce a significant correlation between close predictors (see Figure 1). For each experiment, the support recovery of Δ is evaluated thanks to the so-called F-score given by

$$F = \frac{2\,p_r\,r_e}{p_r + r_e} \quad \text{where} \quad p_r = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad \text{and} \quad r_e = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

are the precision and the recall, respectively, and where T/F and P/N stand for true/false and positive/negative. To assess prediction skills, n_e randomly chosen observations are used for estimation (for different n_e) and the remaining $n_v = n - n_e = 100$ independent observations serve to compute the mean squared prediction error (MSPE). The results are compared to the ones obtained *via* the penalized maximum of likelihood (PML) approach of [33] thanks to the correctly adapted implementations of [6] and [21], with a cross-validated tuning parameter. In addition, we compute the sparse precision matrix estimations given by the graphical Lasso (GLasso) of [10], and by the CLIME algorithm of [4], using the `R` packages `glasso` and `fastclime`, respectively. Note that we always keep a small value for q, so Δ is penalized but not Ω_y when possible (PML and GLasso). Finally, the recent approach of [25], called ANT and based on the individual estimations of the partial correlations, is also implemented. Unlike PML, GLasso and CLIME, sparsity is not the result of penalizations for ANT but, instead, a threshold is deduced from the asymptotic normality of the estimates to decide which are significant and which can be set to zero. Let us add some preliminary

[1]The codes and the dataset are available at `https://github.com/FredericProia/BayesPGGM`

comments about the methods compared in these simulations, all related to high-dimensional precision matrix estimation.

- There is a important advantage in favor of our Bayesian approaches, PML and ANT because they do not need the estimation of $\Omega_x \in \mathbb{S}_{++}^p$. Indeed, extracting the estimation of $\Delta \in \mathbb{R}^{q\times p}$ and $\Omega_y \in \mathbb{S}_{++}^q$ from that of the full precision matrix $\Omega \in \mathbb{S}_{++}^{q+p}$ may generate a drastic bias when $p \gg q$, and that explains in particular why GLasso and CLIME give pretty bad results in what follows.
- In its standard version, ANT is not designed to produce column-sparsity or group-sparsity in Δ. So, by considering multiple testing at the column or even group level, we allow groups of coefficients to be zeroed simultaneously. We have observed that this modified ANT method (called ANT* in the simulations) loses a bit in prediction quality but is greatly improved for support recovery.
- Unfortunately, this is not appropriate for PML, GLasso and CLIME. It is therefore not surprising that they are largely outperformed by our Bayesian models and ANT* for (gs) and (sgs). Using group-penalties, which to the best of our knowledge still does not exist, should improve the results of these methods to some extent.

The seven scenarios below, from Scen. 0 to Scen. 6, as heterogeneous as possible, represent the diversity of the situations (high-dimensionality, kind of sparsity, dimension of the responses, coefficients hard to detect, etc.). We repeat each one $N = 100$ or $N = 50$ times, depending on the computation times involved, and the numerical results for $n_e = 400$ and uncorrelated predictors are summarized in Table 1. In addition, the evolution of MSPE is represented on Figure 1 for Scen. 1, 3 and 5, when n_e grows from 100 to 500, both for uncorrelated and correlated predictors. The three configurations (s), (gs) and (sgs) are tested on the grouped scenarios (from Scen. 3 to Scen. 6) with the adaptative shrinkage.

- *Scenario 0 (small dimension, no sparsity).* Let $q = 1$, $p = 5$ and set $\omega_y = 1$. We fill Δ with $\mathcal{N}(0, 2\,\omega_y)$ coefficients.
- *Scenario 1 (sparse direct links, univariate responses).* Let $q = 1$, $p = 50$ and set $\omega_y = 1$. We randomly choose 10 locations of Δ filled with $\mathcal{N}(0, \omega_y)$ coefficients while the others are zero.
- *Scenario 2 (sparse direct links, multivariate responses).* Let $q = 2$, $p = 80$ and set $\Omega_y = 2\,C_2^{-1}$ with $\rho = 0.5$. We randomly choose 10 columns of Δ filled with $\mathcal{N}_2(0, \Omega_y)$ coefficients while the others are zero.
- *Scenario 3 (group-sparse direct links, univariate responses).* Let $q = 1$, $p = 320$ and set $\omega_y = 1$. We consider $m = 5$ groups of size 100, 10, 100, 10 and 100. The two groups of size 10 are filled with $\mathcal{N}(0, 0.5\,\omega_y)$ and $\mathcal{N}(0, \omega_y)$ coefficients, respectively, while the other groups are zero.
- *Scenario 4 (group-sparse direct links, multivariate responses).* Let $q = 3$, $p = 500$ and set $\Omega_y = 3\,C_3^{-1}$ with $\rho = 0.5$. We divide the columns of Δ into $m = 25$ groups of size 20. We randomly choose 3 groups filled with $\mathcal{N}_3(0, 0.5\,\Omega_y)$, $\mathcal{N}_3(0, \Omega_y)$ and $\mathcal{N}_3(0, 1.5\,\Omega_y)$ coefficients, respectively, while the other groups are zero.
- *Scenario 5 (sparse-group-sparse direct links, univariate responses).* Let $q = 1$, $p = 150$ and set $\omega_y = 1$. We consider $m = 3$ groups of size 50. Only the second group is non-zero, into which we randomly fill 10 locations with $\mathcal{N}(0, \omega_y)$ coefficients.
- *Scenario 6 (sparse-group-sparse direct links, multivariate responses).* Let $q = 5$, $p = 1000$ and set $\Omega_y = 5\,C_5^{-1}$ with $\rho = 0.5$. We divide the columns of Δ into $m = 20$ groups of

size 50, and a randomly chosen one is half filled with $\mathcal{N}_5(0, \Omega_y)$ coefficients. The others columns of Δ are zero.

Scenario 0					
Mod.	Shr.	MSPE	F	p_r	r_e
(s-or)	-	1.01 (0.11)	1.00	1.00	1.00
(s)	(ad)	1.03 (0.13)	1.00	1.00	1.00
(s)	(gl)	1.03 (0.13)	1.00	1.00	1.00
PML	-	1.01 (0.16)	1.00	1.00	1.00
GLasso	-	1.00 (0.15)	1.00	1.00	1.00
CLIME	-	1.00 (0.15)	1.00	1.00	1.00
ANT*	-	1.04 (0.13)	1.00	1.00	1.00
Hyperparam. $\pi = 0$					

Scenario 1					
Mod.	Shr.	MSPE	F	p_r	r_e
(s-or)	-	1.02 (0.13)	0.95	1.00	0.90
(s)	(ad)	1.04 (0.13)	0.95	1.00	0.90
(s)	(gl)	1.03 (0.13)	0.95	1.00	0.90
PML	-	1.08 (0.15)	0.82	0.69	1.00
GLasso	-	2.37 (0.96)	0.78	0.77	0.80
CLIME	-	2.52 (0.98)	0.79	0.78	0.80
ANT*	-	1.25 (0.22)	0.87	0.85	0.90
Hyperparam. $(25, 1)$					

Scenario 2					
Mod.	Shr.	MSPE	F	p_r	r_e
(s-or)	-	0.52 (0.09)	0.95	1.00	0.90
(s)	(ad)	0.54 (0.09)	0.95	1.00	0.90
(s)	(gl)	0.55 (0.08)	0.95	1.00	0.90
PML	-	0.77 (0.15)	0.86	1.00	0.75
GLasso	-	1.74 (0.49)	0.72	0.91	0.60
CLIME	-	1.11 (0.35)	0.73	0.76	0.70
ANT*	-	1.04 (0.44)	0.90	0.89	0.91
Hyperparam. $(80, 1)$					

Scenario 3					
Mod.	Shr.	MSPE	F	p_r	r_e
(gs-or)	-	1.03 (0.27)	1.00	1.00	1.00
(gs)	(ad)	1.04 (0.27)	1.00	1.00	1.00
(gs)	(gl)	1.04 (0.34)	1.00	1.00	1.00
(s)	(ad)	1.16 (0.27)	0.92	1.00	0.85
(sgs)	(ad)	1.07 (0.25)	0.92	1.00	0.86
PML	-	1.80 (0.36)	0.89	1.00	0.80
GLasso	-	4.23 (1.61)	0.58	0.50	0.70
CLIME	-	2.98 (1.22)	0.68	0.90	0.55
ANT*	-	1.52 (0.95)	1.00	1.00	1.00
Hyperparam. $(100, 1) - (5, 1) - (5, 1, 25, 1)$					

Scenario 4					
Mod.	Shr.	MSPE	F	p_r	r_e
(gs-or)	-	0.40 (0.14)	1.00	1.00	1.00
(gs)	(ad)	0.45 (0.16)	1.00	1.00	1.00
(gs)	(gl)	0.46 (0.17)	1.00	1.00	1.00
(s)	(ad)	0.52 (0.18)	0.98	1.00	0.96
(sgs)	(ad)	0.48 (0.17)	0.99	1.00	0.98
PML	-	3.18 (0.53)	0.75	0.94	0.62
GLasso	-	9.46 (1.38)	0.46	0.66	0.35
CLIME	-	8.32 (1.51)	0.48	0.45	0.52
ANT*	-	6.53 (1.22)	1.00	1.00	1.00
Hyperparam. $(100, 1) - (25, 1) - (50, 1, 50, 1)$					

Now, let us try to summarize our observations. In terms of support recovery, the Bayesian spike-and-slab framework and the modified ANT* method give results incomparably better than the sparsity-inducing penalized approaches (PML, GLasso and CLIME). As suggested in Rem. 3.3 of [21], this may be a consequence of the fact that the cross-validation steps calibrate the models to reach the best prediction error, sometimes at the cost of support recovery by picking a small penalty level. The superiority of ANT over GLasso and CLIME is recognized and discussed in [25], but this also highlights the ability of our Bayesian models to reach good results both in prediction and in support recovery. It can also be seen that (s) gives weaker results than (sgs) in the grouped scenarios, probably due to the fact that it does not take into account the group structure, but still better than the penalized methods. However, the computational times involved (see remarks below) make (s) less relevant than (sgs) in these situations, even if the results are not drastically different. Unsurprisingly,

Scenario 5					
Mod.	Shr.	MSPE	F	p_r	r_e
(sgs-or)	-	1.00 (0.15)	0.96	1.00	0.92
(sgs)	(ad)	1.04 (0.16)	0.95	1.00	0.91
(sgs)	(gl)	1.03 (0.16)	0.91	1.00	0.84
(s)	(ad)	1.08 (0.14)	0.93	1.00	0.87
(gs)	(ad)	1.24 (0.19)	0.33	0.20	1.00
PML	-	1.92 (0.60)	0.89	1.00	0.80
GLasso	-	3.48 (1.30)	0.78	0.86	0.71
CLIME	-	1.88 (0.92)	0.79	1.00	0.65
ANT*	-	1.26 (0.98)	0.88	0.86	0.90
Hyperparam. $(50, 1) - (3, 1) - (3, 1, 50, 1)$					

Scenario 6					
Mod.	Shr.	MSPE	F	p_r	r_e
(sgs-or)	-	0.21 (0.13)	1.00	1.00	1.00
(sgs)	(ad)	0.24 (0.32)	1.00	1.00	1.00
(sgs)	(gl)	0.24 (0.33)	1.00	1.00	1.00
(s)	(ad)	0.29 (0.26)	0.98	1.00	0.96
(gs)	(ad)	0.31 (0.30)	0.67	0.50	1.00
PML	-	0.50 (0.17)	0.83	0.95	0.74
GLasso	-	3.83 (0.77)	0.50	0.97	0.34
CLIME	-	2.98 (0.51)	0.51	1.00	0.34
ANT*	-	2.10 (0.72)	1.00	1.00	1.00
Hyperparam. $(100, 1) - (20, 1) - (20, 1, 50, 1)$					

TABLE 1. Medians of the mean squared prediction errors (with standard deviations), F-scores, precisions and recalls after $N = 100$ repetitions of Scen. 0 to Scen. 6 ($N = 50$ for Scen. 4 and Scen. 6), with $n_e = 400$ and uncorrelated predictors. The suffix -or is used to denote 'oracle' settings. The hyperparameters chosen for the prior spike probability are indicated in the last row of each table, from left to right: (a, b) for (s) and (gs), (a_1, b_1, a_2, b_2) for (sgs).

(gs) is not suitable in the sparse-group-sparse settings in terms of support recovery. Our experiments show that it is able to identify influential groups without being mistaken but, even though the resulting estimates are small where they should be zero, it is not designed to be used for bi-level selection. Figure 1 shows that the results are pretty stable from $n_e = 200$ observations in the learning set: for $n_e < 200$ the MSPEs are rather chaotic before stabilizing. The same figure also highlights that the presence of correlation in the predictors does not seem to have a significant effect on the estimation procedure, except for small size samples and high correlation where the degradation is noticeable. Overall, the real strength of the Bayesian spike-and-slab approach is clearly the support recovery of the direct links between predictors and responses but it seems that one can hardly expect to deal with very high-dimensional studies as long as we do not impose a group structure or a huge degree of sparsity. The highly competitive MSPEs obtained confirm the relevance of Bayesian PGGMs not only for variable selection but also for prediction purposes in the context of high-dimensional regressions.

6.2. **Identification of a sparse set of predictors in a real dataset.** Let us now study the `Hopx` dataset, fully described in [23]. It contains $p = 770$ genetic markers spread over $m = 20$ chromosomes from $n = 29$ inbred rats. It also contains the corresponding measured gene expression levels of $q = 4$ tissues (adrenal gland, fat, heart and kidney). The goal is to identify a sparse set of predictors that jointly explain the outcomes, with the natural group structure formed by chromosomes (see Table 2). This dataset has already been analyzed in [16], using a Bayesian regression without group structure, and later in [17] including group and sparse-group structures. So the PGGM is supposed to bring new perspectives about relationships in terms of partial correlations. A particularity of this dataset is that the responses are very correlated, so we should expect an estimation of Ω_y^{-1} with significant non-diagonal elements and a clear advantage in using PGGMs. Indeed, a predictor considered to be influencing all the outcomes could be the result of a direct relation to one tissue propagated to the others by an artificial correlation effect. As can be seen on Figure 2, the

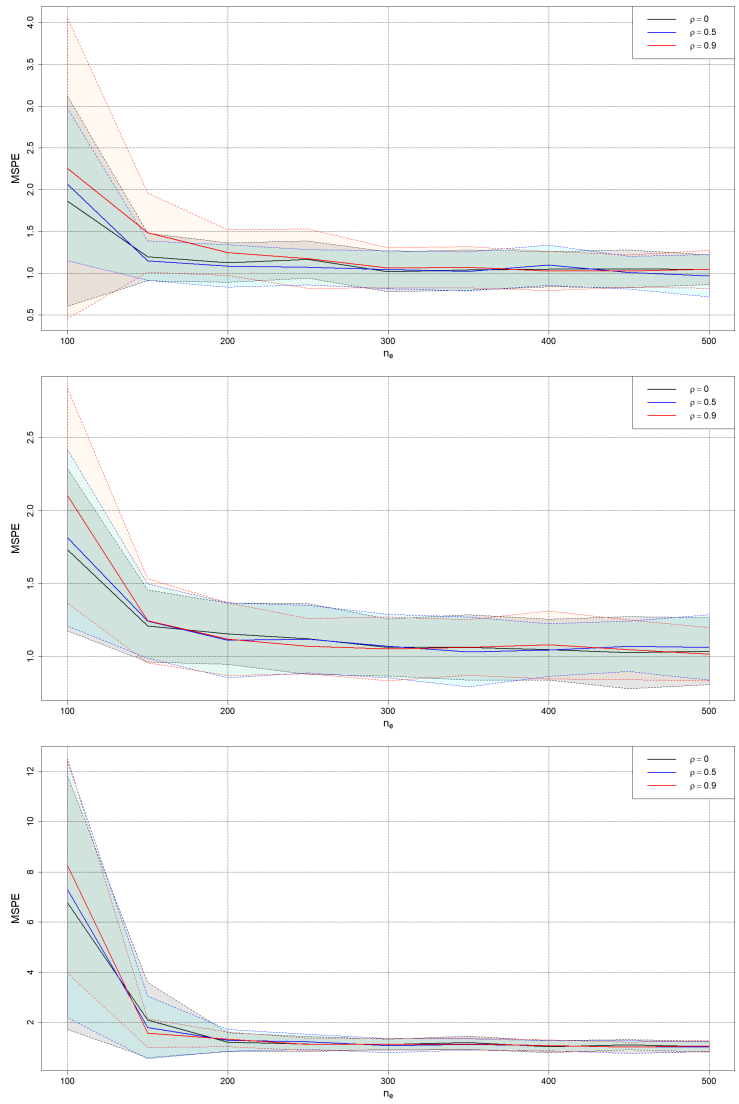


FIGURE 1. Medians of the mean squared prediction errors obtained after $N = 100$ repetitions of Scen. 1 (top), Scen. 3 (middle) and Scen. 5 (bottom) with ± 1 standard deviation and n_e growing from 100 to 500. The black curves correspond to uncorrelated predictors ($\rho = 0$) while the blue and red curves correspond to correlated predictors ($\rho = 0.5$ and $\rho = 0.9$, respectively).

predictors are also highly correlated with their neighbors (for the sake of readability, we only represent the correlogram of predictors located on chromosomes 8, 9 and 10).

Chr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nb.	74	67	63	60	39	45	52	43	31	51	21	26	33	22	15	27	18	30	34	19

TABLE 2. Number of markers on each chromosome, which correspond to the sizes κ_g of each group for $1 \leqslant g \leqslant 20$ when running (gs) and (sgs).

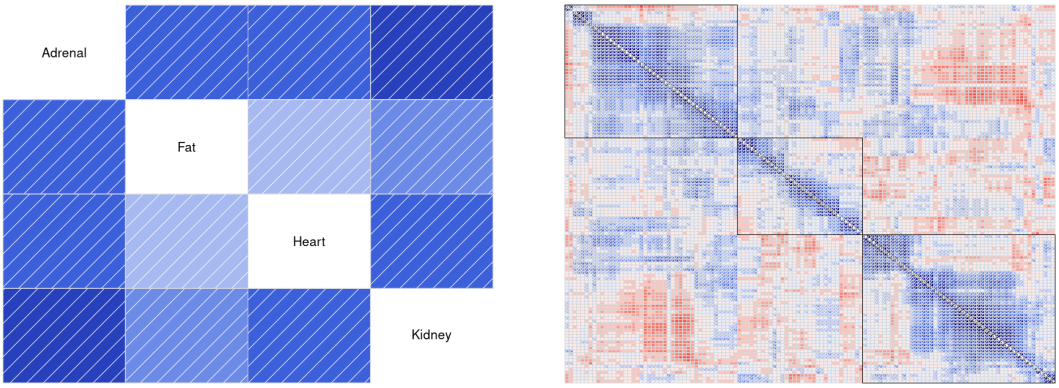


FIGURE 2. Correlogram of responses (left) and correlogram of predictors located on chromosomes 8, 9 and 10 (right). The colormap associates red with negative correlations and blue with positive correlations.

The small sample size relative to the number of covariates (29/770) weakens the study. To strengthen our conclusions, we decided to run $N = 100$ experiments based on 25 randomly chosen observations and to aggregate the results. We first investigate the selection of predictors at the chromosomes scale, *i.e.* we run (gs) according to the previous protocol with an adaptative shrinkage and we choose $(a, b) = (1, 20)$ in the prior probability π. The empirical distribution of the posterior probability of inclusion for each chromosome is represented on the left of Figure 3. The selection procedure focuses on chromosomes 14, 15 and 17 (and not just on chromosomes 2 and 3 as in [17]) but the estimation process gives an overwhelming advantage to chromosome 14, far ahead of its neighbors. This is undoubtedly the influence of `D14Mit3`, a marker located on chromosome 14 and known to have a very significant effect on this dataset. The main conclusion to be drawn at this stage is that chromosome 14 has a positive effect on `Fat` and a *negative* effect on `Heart`, as can also be seen on the right of Figure 3. Therefore, it is likely that the overall positive influence of `D14Mit3` identified by previous authors is due to the combination of a direct positive link with `Fat`, a direct negative link with `Heart` and a correlation effect from the outcomes. This hypothesis is given additional credibility by the numerical results: from (gs), the corresponding column of Δ is approximately $(0.00, 0.04, -0.09, 0.00)$ which, through relations (1.1), leads to $(0.15, 0.25, 0.34, 0.21)$ as estimated regression coefficients. This roughly corresponds to the values indicated in Tab. 2 of [17], at least for the main effect on `Heart`. Thus for chromosome 14, the numerical results coincide but the interpretations are clearly different. Of course, similar reasonings can be carried out for the less influent chromosomes.

It is perhaps more interesting to look for a bi-level selection in order to identify a sparse set of markers and not only chromosomes. In this regard, (sgs) is launched using the same statistical protocol, adaptative shrinkage and hardly informative hypermarameters $a_1 = 3$, $b_1 = 1$, $a_2 = 1$ and $b_2 = 1$ which happen to be sufficient to generate a huge degree of sparsity. While many chromosomes are excluded from the model given by (gs), with (sgs) we see some contributions localized in certain chromosomes having little influence when taken as a whole. At the markers scale, the randomness of the sampler and the high level of correlation

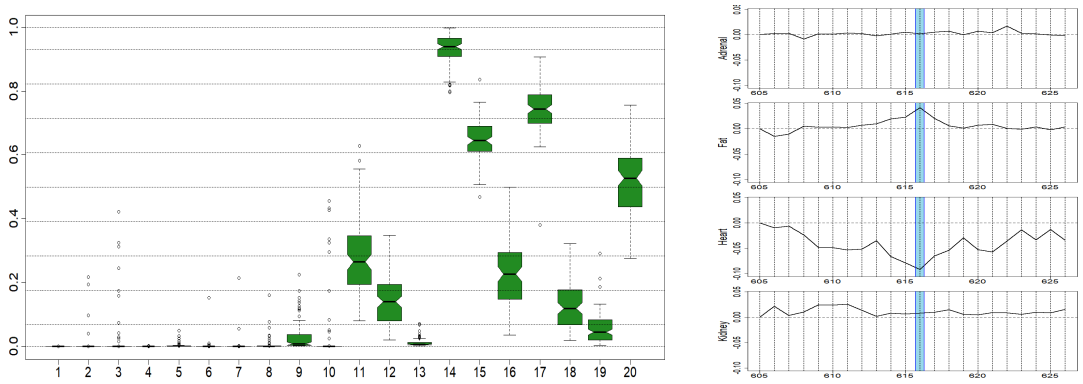


FIGURE 3. Empirical distribution of the posterior probability of inclusion estimated by (gs) for each chromosome (left). Aggregated (gs) estimation of Δ on chromosome 14 with `D14Mit3` hilighted (right).

between close predictors probably explain the presence of artifacts which sometimes make it difficult to distinguish the real contributions from the background noise. We therefore use the $N = 100$ experiments to build 95% confidence intervals and keep only significant estimates. By way of example, Figure 4 displays the results obtained on chromosomes 7, 8 and 14. The main markers standing out are summarized in Table 3 together with the kind of direct influences detected. Markers already highlighted in [16] or [17] are also indicated. One can see that most of our conclusions coincide, but new markers are suggested (especially on chromosome 8) and others have disappeared. Overall, the more stringent statistical protocol that we used led to the retention of fewer predictors with more guarantee. An important consequence of this study is the new interpretations in terms of direct influences allowed by PGGMs. Especially as the residual correlations, hidden in the estimation of $R = \Omega_y^{-1}$ and closely related to the correlations between the responses, are very high (greater than 0.7), as we suspected from Figure 2.

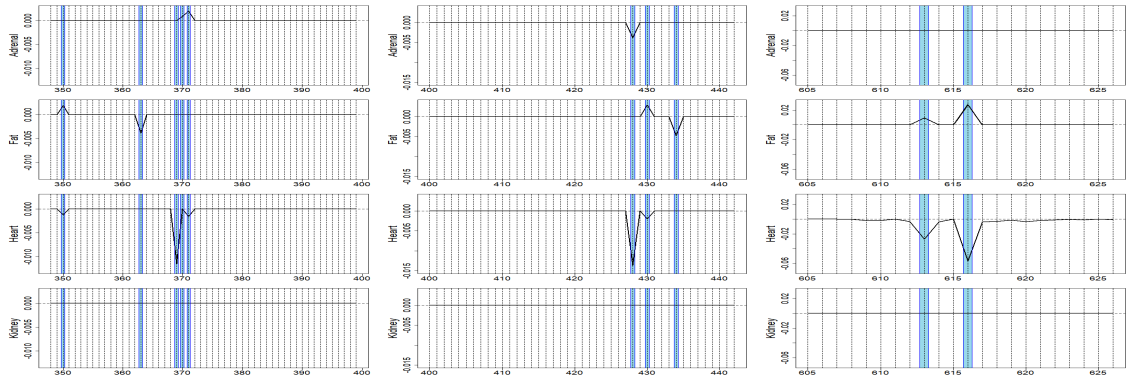


FIGURE 4. Aggregated (sgs) estimation of Δ on chromosomes 7, 8 and 14, from left to right. The hilighted markers are `D7Cebr205s3`, `D7Mit6`, `D7Rat19`, `Myc` and `D7Rat17` for chromosome 7, `D8Mgh4`, `D8Rat135` and `Rbp2` for chromosome 8 and `D14Rat8` and `D14Mit3` for chromosome 14.

Chromosomes	Markers	Main direct influences
3	D3Mit16*	Adrenal+ Heart–
7	D7Cebr205s3*	Fat+ Heart–
	D7Mit6*	Fat–
	D7Rat19*	Heart–
	Myc*	Adrenal+
	D7Rat17	Adrenal+ Heart–
8	D8Mgh4	Adrenal– Heart–
	D8Rat135	Fat+ Heart–
	Rbp2	Fat–
10	D10Rat33*	Adrenal+
	D10Mit3*	Adrenal+
	D10Rat31*	Fat–
11	D11Rat47	Fat–
14	D14Rat8*	Fat+ Heart–
	D14Mit3*	Fat+ Heart–
15	D15Cebr7s13	Kidney–
	D15Rat21*	Adrenal+ Kidney–
17	Prl	Adrenal– Kidney–
20	D20Rat55	Kidney–

TABLE 3. Main relations detected by (sgs). X* means that marker X has already been suggested by previous authors in this dataset. Y– (Y+) means that response Y is negatively (positively) influenced by X.

6.3. **Discussion and Conclusion.** To conclude, we would like to draw the attention of the reader to some weaknesses of the study, still under investigation. On the one hand, as soon as p is large (say, $p \geqslant 500$), the Bayesian studies should be conducted with a group structure or by promoting very sparse settings because due to the outline of the sampler, looping over each column of Δ may quickly become intractable. A group structure limits the number of loops (only $m \ll p$ per sampler iteration), although each loop may require the generation of large Gaussian vectors (up to $(q \times \kappa_g)$-dimensional), so compromises are needed. Subdividing the dataset is natural when it is intrinsically equipped with a group structure (*e.g.* that of the previous section), we could suggest otherwise a clustering of the set of predictors to gather similar entries and control the size of the groups. At this stage, our procedures cannot compete with the Lasso-type algorithms (GLasso, CLIME or even ANT) in terms of computational times. This is an issue on which future studies should focus (ongoing works are devoted to translating the samplers into more efficient environments), enhanced MCMC methods may also be useful or novel computational strategies like the 'shotgun' stochastic algorithm of [31]. On the other hand, the procedures are obviously very sensitive to the initialization of the sampler, especially when $p \gg n$. For example, the term $|I_{\kappa_g} + \lambda_g \mathbb{X}_g^t \mathbb{X}_g|$ is likely to explode when κ_g is large and $\lambda_g > 1$, that is why λ_g has to be carefully controlled *via* an accurate initial choice of ℓ_g. Our heuristic approach is to initialize ℓ_g such that $\mathbb{E}[\lambda_g] < 1$ to control the behavior of $|I_{\kappa_g} + \lambda_g \mathbb{X}_g^t \mathbb{X}_g|$ during the first iterations. This works pretty well in practice, but needs to be done on a case-by-case basis, which could be improved. From a theoretical point of view, we should obviously enhance the estimation procedure by sampling from the $\mathcal{MGIG}_q$ distribution for $q > 1$, and not using the mode. Our fallback solution gives satisfactory but not completely rigorous results. In addition, it could be interesting to generalize the support recovery guarantee of Proposition 3.2 to (sgs),

which is certainly possible at the cost of a few additional developments. Overall, our study shows that for the moderate values of p (up to 10^3 or 10^4), the Bayesian approach of the partial Gaussian graphical models is a very serious alternative to the frequentist penalized estimations, for prediction but also and especially for support recovery.

Acknowledgements and Fundings. The authors thank ALM (Angers Loire Métropole) and the ICO (Institut de Cancérologie de l'Ouest) for the financial support. This work is partially financed through the ALM grant and the "Programme opérationnel régional FEDER-FSE Pays de la Loire 2014-2020" noPL0015129 (EPICURE). The authors also thank Mario Campone (project leader and director of the ICO), Mathilde Colombié (scientific coordinator of EPICURE clinical trial) and Fadwa Ben Azzouz, biomathematician in Bioinfomics, for the initiation, the coordination and the smooth running of the project.

References

[1] Bai, R., Moran, G. E., Antonelli, J. L., Chen, Y., and Boland, M. R. Spike-and-slab group lassos for grouped regression and sparse generalized additive models. *J. Am. Stat. Assoc.* (2020), 1–14.

[2] Banerjee, O., El Ghaoui, L., and d'Aspremont, A. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res. 9* (2008), 485–516.

[3] Brown, P. J., Vannucci, M., and Fearn, T. Multivariate Bayesian variable selection and prediction. *J. R. Statist. Soc. B. 60*, 3 (1998), 627–641.

[4] Cai, T., Liu, W., and Luo, X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Am. Stat. Assoc. 106*, 494 (2011), 594–607.

[5] Cai, T., and Zhou, H. Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Stat. 40*, 5 (2012), 2389–2420.

[6] Chiquet, J., Mary-Huard, T., and Robin, S. Structured regularization for conditional Gaussian graphical models. *Stat. Comput. 27*, 3 (2017), 789–804.

[7] Eltoft, T., Kim, T., and Lee, T. Multivariate scale mixture of Gaussians modeling. In *Independent Component Analysis and Blind Signal Separation* (2006), Springer Berlin Heidelberg, pp. 799–806.

[8] Fang, Y., Karlis, D., and Subedi, S. A Bayesian approach for clustering skewed data using mixtures of multivariate normal-inverse Gaussian distributions. *arXiv:2005.02585* (2020).

[9] Fazayeli, F., and Banerjee, A. *The Matrix Generalized Inverse Gaussian distribution: properties and applications*, vol. 9851 of *Frasconi P., Landwehr N., Manco G., Vreeken J. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2016. Lecture Notes in Computer Science.* Springer, Cham., 2016.

[10] Friedman, J., Hastie, T., and Tibshirani, R. Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics. 9*, 3 (2008), 432–441.

[11] Gan, L., Yang, X., Narisetty, N., and Liang, F. Bayesian joint estimation of multiple graphical models. In *Advances in Neural Information Processing Systems* (2019), vol. 32, Curran Associates, Inc.

[12] Giraud, C. *Introduction to High-Dimensional Statistics.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014.

[13] Hastie, T., Tibshirani, R., and Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations.* Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC Press, 2015.

[14] Li, Y., Nan, B., and Zhu, J. Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. *Biometrics. 71* (2015), 354–363.

[15] Li, Z., Mccormick, T., and Clark, S. Bayesian joint spike-and-slab graphical Lasso. In *Proceedings of the 36th International Conference on Machine Learning* (2019), vol. 97 of *Proceedings of Machine Learning Research*, PMLR, pp. 3877–3885.

[16] Liquet, B., Bottolo, L., Campanella, G., Richardson, S., and Chadeau-Hyam, M. R2GUESS: A graphics processing unit-based R package for Bayesian variable selection regression of multivariate responses. *J. Stat. Softw. 69*, 2 (2016), 1–32.

[17] Liquet, B., Mengersen, K., Pettitt, A. N., and Sutton, M. Bayesian variable selection regression of multivariate responses for group data. *Bayesian Anal. 12*, 4 (2017), 1039–1067.
[18] Maathuis, M., Drton, M., Lauritzen, S. L., and Wainwright, M. *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2018.
[19] Massam, H., and Wesołowski, J. The Matsumoto-Yor property and the structure of the Wishart distribution. *J. Multivariate. Anal. 97* (2006), 103–123.
[20] Meinshausen, N., and Bühlmann, P. High-dimensional graphs and variable selection with the Lasso. *Ann. Stat. 34*, 3 (2006), 1436–1462.
[21] Okome Obiang, E., Jézéquel, P., and Proïa, F. A partial graphical model with a structural prior on the direct links between predictors and responses. *ESAIM Probab. Stat. 25* (2021), 298–324.
[22] Park, T., and Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc. 103*, 482 (2008), 681–686.
[23] Petretto, E., Bottolo, L., Langley, S. R., Heinig, M., McDermott-Roe, C., Sarwar, R., Pravenec, M., Hübner, N., Aitman, T. J., Cook, S. A., and Richardson, S. New insights into the genetic control of gene expression using a bayesian multi-tissue approach. *PLOS Comput. Biol. 6*, 4 (2010), 1–13.
[24] Ravikumar, P., Wainwright, M., Raskutti, G., and Yu, B. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat. 5* (2011), 935–980.
[25] Ren, Z., Sun, T., Zhang, C. H., and Zhou, H. H. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Stat. 43*, 3 (2015), 991–1026.
[26] Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. Sparse permutation invariant covariance estimation. *Electron. J. Stat. 2* (2008), 494–515.
[27] Sohn, K. A., and Kim, S. Joint estimation of structured sparsity and output structure in multiple-output regression via inverse-covariance regularization. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics.* (2012), vol. 22 of *Proceedings of Machine Learning Research*, PMLR, pp. 1081–1089.
[28] Wei, R., Reich, B. J., Hoppin, J. A., and Ghosal, S. Sparse Bayesian additive nonparametric regression with application to health effects of pesticides mixtures. *Statist. Sinica 30* (2020), 55–79.
[29] Xu, X., and Ghosh, M. Bayesian variable selection and estimation for Group Lasso. *Bayesian Anal. 10*, 4 (2015), 909–936.
[30] Xu, Z., Schmidt, D. F., Makalic, E., Qian, G., and Hopper, J. L. Bayesian grouped horseshoe regression with application to additive models. In *AI 2016: Advances in Artificial Intelligence* (2016), Springer International Publishing, pp. 229–240.
[31] Yang, X., and Narisetty, N. Consistent group selection with bayesian high dimensional modeling. *Bayesian Anal. 15*, 3 (2020), 909–935.
[32] Yuan, M., and Lin, Y. Model selection and estimation in the Gaussian graphical model. *Biometrika. 94*, 1 (2007), 19–35.
[33] Yuan, X. T., and Zhang, T. Partial Gaussian graphical model estimation. *IEEE. T. Inform. Theory. 60*, 3 (2014), 1673–1687.

Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France.
Email address: okome@math.univ-angers.fr

1 Unité de Bioinfomique, Institut de Cancérologie de l'Ouest, Bd Jacques Monod, 44805 Saint Herblain Cedex, France.
2 SIRIC ILIAD, Nantes, Angers, France.
3 CRCINA, INSERM, CNRS, Université de Nantes, Université d'Angers, Institut de Recherche en Santé-Université de Nantes, 8 Quai Moncousu - BP 70721, 44007, Nantes Cedex 1, France.
Email address: pascal.jezequel@ico.unicancer.fr

Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France.
Email address: frederic.proia@univ-angers.fr

Bibliographie

Chiquet, J., Mary-Huard, T. et Robin, S. (2017). Structured regularization for conditional Gaussian graphical models. *Stat. Comput.*, 27(3):789–804.

Friedman, J., Hastie, T. et Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics.*, 9(3):432–441.

Giraud, C. (2014). *Introduction to High-Dimensional Statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Hastie, T., Tibshirani, R. et Wainwright, M. (2015). *Statistical Learning with Sparsity : The Lasso and Generalizations*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC Press.

Liquet, B., Mengersen, K., Pettitt, A. N. et Sutton, M. (2017). Bayesian variable selection regression of multivariate responses for group data. *Bayesian Anal.*, 12(4):1039–1067.

Maathuis, M., Drton, M., Lauritzen, S. L. et Wainwright, M. (2018). *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press.

Massam, H. et Wesołowski, J. (2006). The Matsumoto-Yor property and the structure of the Wishart distribution. *J. Multivariate. Anal.*, 97:103–123.

Okome Obiang, E., Jézéquel, P. et Proïa, F. (2021). A partial graphical model with a structural prior on the direct links between predictors and responses. *ESAIM Probab. Stat.*, 25:298–324.

Okome Obiang, E., Jézéquel, P. et Proïa, F. (2022). A Bayesian approach for partial Gaussian graphical models with sparsity. *To appear in Bayesian Anal.*

Xu, X. et Ghosh, M. (2015). Bayesian variable selection and estimation for Group Lasso. *Bayesian Anal.*, 10(4):909–936.

Yang, X. et Narisetty, N. (2020). Consistent group selection with Bayesian high dimensional modeling. *Bayesian Anal.*, 15(3):909–935.

Yuan, X. T. et Zhang, T. (2014). Partial Gaussian graphical model estimation. *IEEE. T. Inform. Theory.*, 60(3):1673–1687.

Chapitre 3

Applications aux sciences du vivant

Dans le cadre de collaborations avec l'Institut de Recherche en Horticulture et Semences (IRHS), unité mixte de recherche INRAE/Université d'Angers/Institut Agro, des études sont régulièrement menées afin de répondre à des besoins interdisciplinaires en traitement et analyse des données. Des travaux ont été valorisés à ce jour, nous allons dans ce dernier chapitre nous contenter de résumer l'un d'entre eux avant de nous focaliser sur un autre. Tous deux traitent de populations de rosiers, du point de vue phénotypique (courbes de floraison) et génotypique (reconstruction de généalogies).

3.1 Modélisation de courbes de floraison

L'article Proïa *et al.* (2016) publié dans *Journal of Theoretical Biology* est issu d'un travail en commun avec A. Pernet, T. Thouroude, G. Michel et J. Clotault, de l'IRHS. Il est dédié à l'étude de caractères phénotypiques dans une population végétale à des fins de modélisation et de classification.

Résumé

On souhaite modéliser les courbes de floraison d'une population de rosiers (mesurées en densité de fleurs sur la plante au cours du temps) avec un double objectif : mettre en évidence et décrire par des indicateurs pertinents les vagues de floraison, puis en déduire une classification des rosiers. Nous proposons à cet égard de sélectionner pour chaque individu un modèle de mélange gaussien estimé sur un échantillon dont la distribution correspond à la courbe (en tenant compte des valeurs manquantes), voir par exemple la Figure 3.1 ci-dessous. La sélection du nombre de composantes dans les mélanges se fait par minimisation du critère *ad hoc* défini par

$$\forall\, k \geqslant 1, \quad \mathrm{BIC}^*(k) = \big(c + \mathrm{BIC}(k)\big)\left(1 + \mathrm{e}^{-\alpha\, d_k}\right) \tag{3.1}$$

avec $c \geqslant 0$, $\alpha \geqslant 0$, $d_1 = +\infty$ et pour $k \geqslant 2$,

$$d_k = \min_{\substack{1 \leqslant j_1,\, j_2 \leqslant k \\ (j_1 \neq j_2)}} |\,\widetilde{\mu}_{j_1} - \widetilde{\mu}_{j_2}| \tag{3.2}$$

où $\widetilde{\mu}_j$ désigne la moyenne estimée de la j-ème composante alors que c et α sont des paramètres de régulation. On pénalise ainsi les vagues trop proches l'une de l'autre,

conséquence de la présence d'asymétrie qui conduit le BIC usuel à suggérer beaucoup trop de composantes. Ce phénomène est particulièrement visible sur la Figure 3.2 où l'on voit que le BIC impose un nombre irréaliste de vagues alors que le BIC* est minimisé pour $k = 3$ composantes.

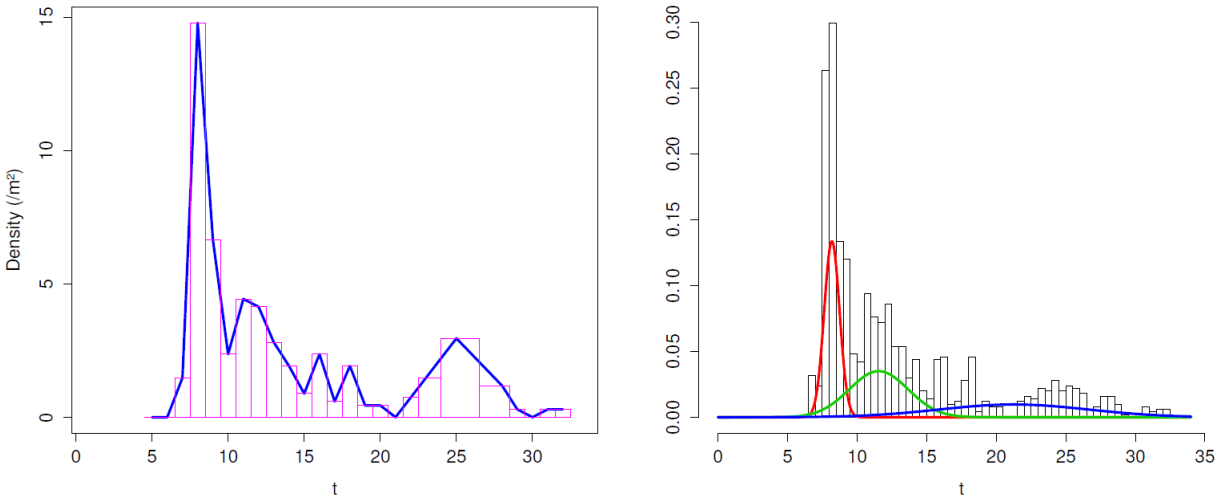


FIGURE 3.1 – Exemple de courbe de floraison (à gauche) modélisée par un mélange gaussien (à droite) à $k = 3$ composantes.

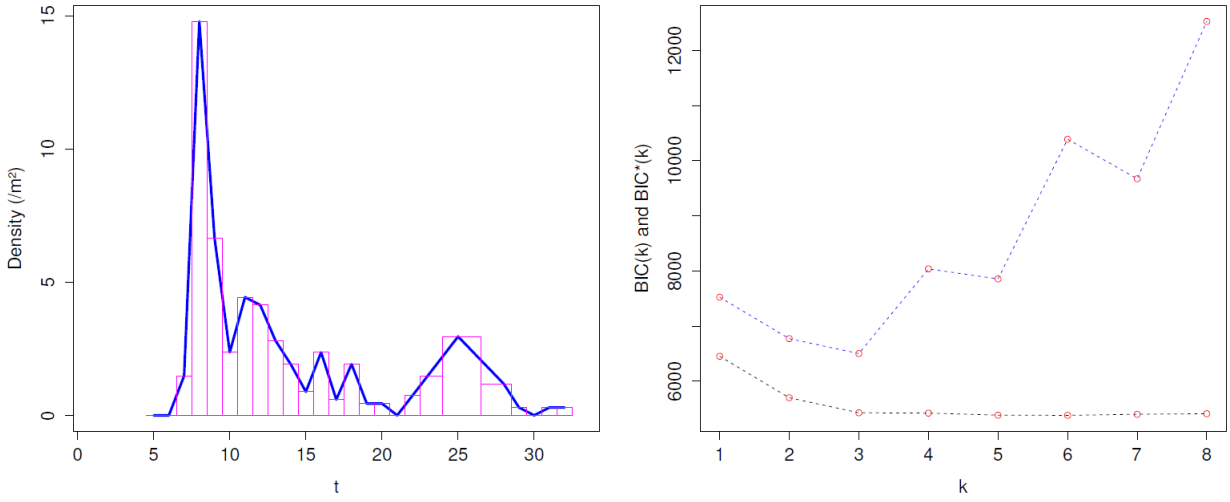


FIGURE 3.2 – Exemple identique au précédent (à gauche) avec évolution des critères BIC en noir et BIC* en bleu (à droite).

De tous ces mélanges, on tire des indicateurs qui permettent de réduire la dimension de l'étude : la courbe de floraison d'un individu se trouve décrite par le nombre de vagues, l'intensité maximale, l'aire sous la courbe de la première floraison, la proportion de l'aire totale correspondant à la première floraison ainsi que son démarrage (la précocité), et quelques autres dont l'énumération serait inutile ici, tous motivés par des arguments issus de l'expertise biologique. Une analyse descriptive de ces indicateurs à travers une ACP

révèle que les traits principaux caractérisant la floraison d'un rosier sont liés à l'intensité de sa remontée de floraison, c'est-à-dire à sa (re-)floraison après la première poussée, et de manière orthogonale à sa précocité. Cela donne finalement lieu à un clustering à deux niveaux : par *k-means* sur le premier plan factoriel et par la méthode des *k-means* longitudinaux (KML) de Genolini et Falissard (2010) sur les courbes renormalisées dans chaque cluster, afin d'obtenir des profils caractéristiques de floraison. La Figure 3.3 montre ainsi les profils obtenus dans le cluster des rosiers non-remontants. En lien avec nos conclusions préalables, on peut constater que la précocité apparaît comme le principal élément discriminant dans une classe où la remontée est inexistante.

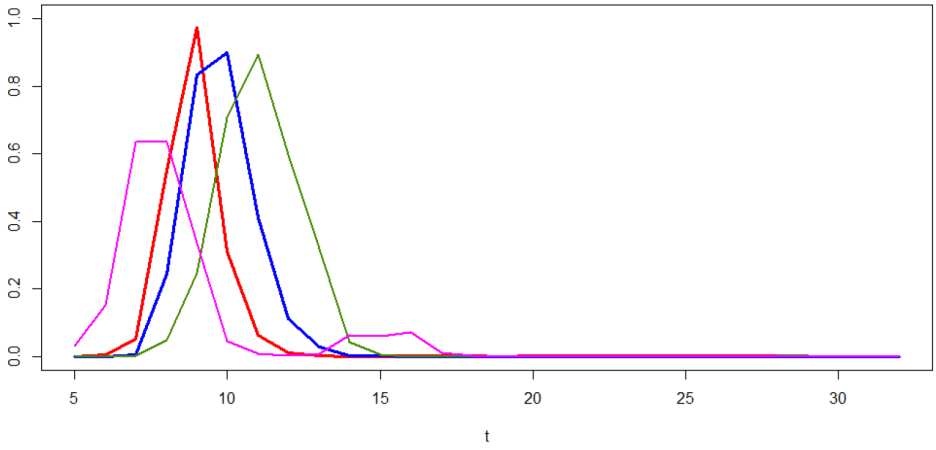


FIGURE 3.3 – Profils issus de KML dans le cluster des rosiers non-remontants.

Perspectives

Les perspectives empiriques de cette étude sont entre les mains des biologistes, qui espèrent que les indicateurs obtenus et les analyses statistiques, croisés avec les conditions spécifiques à l'entretien de cette population de rosiers, permettront de faire avancer d'un pas la compréhension des contrôles génétiques et environnementaux des processus biologiques qui sous-tendent les vagues de floraison. Mais revenons simplement sur un point théorique. L'introduction du critère BIC* pour pallier la dissymétrie des vagues donne de très bons résultats numériques et nous permet d'atteindre les objectifs souhaités, mais elle manque de rigueur de par son côté arbitraire. Avec le recul, il pourrait être judicieux d'aborder la modélisation des courbes par des mélanges Gamma, à l'image de ce que proposent Wiper *et al.* (2001), munis d'un critère de sélection plus conventionnel. D'un point de vue pratique, nos indicateurs sortis sur trois années d'étude sont actuellement utilisés dans le cadre d'un postdoctorat sur de la génétique d'association (GWAS, Genome-Wide Association Studies) : construction de modèles qui essayent d'expliquer la variation d'un phénotype des rosiers de la collection Loubert (celle, angevine, dont l'analyse a été fournie), à partir de la variation à 68000 SNPs (marqueurs moléculaires constitués d'un nucléotide seulement), de la connaissance de la structuration et de "l'apparentement" (qui n'est pas une généalogie, mais une analyse brute relative à la part d'allèles en communs). Le travail s'avère important pour avoir des caractères quantitatifs et pas seulement des classes arbitrairement créées.

3.2 Reconstruction probabiliste de généalogies

Nous en arrivons à la dernière étude que nous souhaitions mettre en avant dans ce mémoire. Proïa *et al.* (2019) est le fruit d'une collaboration avec F. Panloup, C. Trabelsi et J. Clotault, également publiée dans *Journal of Theoretical Biology*. Contrairement à la précédente, cette étude n'est pas une analyse statistique mais une construction probabiliste. À partir de marqueurs génétiques et d'autres informations de nature descriptive, on va chercher à reconstruire rétrospectivement l'arbre généalogique d'une population de rosiers avec comme inspiration les travaux de Chaumont *et al.* (2017), mais dans un contexte plus général.

Résumé

On dispose de marqueurs génétiques issus d'une population de rosiers diploïdes 2x (les chromosomes vont par 2), triploïdes 3x (par 3) et tétraploïdes 4x (par 4), qualificatifs auxquels on en profite pour adjoindre les haploïdes 1x (par 1). Connaissant la date d'obtention de chaque individu, on se propose de mettre au point la construction d'un graphe orienté et probabilisé illustrant les relations de parenté directe les plus probables entre les individus, possédant donc la structure d'arbre. Formellement, on va considérer qu'une généalogie sur notre population $\mathcal{P}$ est un élément de l'ensemble

$$\Upsilon(\mathcal{P}) = \prod_{e \in \mathcal{P}} \big\{ \mathbb{T}(e) \cup (e, \varnothing) \big\} \quad \text{avec} \quad \mathbb{T}(e) = \bigcup_{s \in \mathcal{S}(e)} (e, s) \tag{3.3}$$

où $\mathcal{S}(e) \subset \mathcal{P}^2$ contient les couples (p_1, p_2) avec $p_1 \neq p_2$ génétiquement et chronologiquement candidats à la parenté directe de l'individu e. En conséquence, la vraisemblance d'une généalogie $\mathcal{T} \in \Upsilon(\mathcal{P})$ sera naturellement définie comme la probabilité qu'elle a d'être observée,

$$\ell(\mathcal{T}) = \mathbb{P}(\mathcal{T}) \tag{3.4}$$

grâce au support d'un ensemble d'hypothèses et de règles de calcul formalisant notre modèle. La première difficulté est liée aux schémas de reproduction qui sont plus complexes que le schéma standard $\{a, b\} \times \{c, d\} \mapsto \{ac, ad, bc, bd\}$ valable en présence de diploïdie. Le mix 2x/3x/4x donne lieu à 10 schémas de reproduction, chacun d'entre eux engendré par la production des gamètes spécifiques à chaque ploïdie : on retiendra ici qu'un 2x peut produire 2 gamètes haploïdes, qu'un 3x peut produire 3 gamètes haploïdes et 3 gamètes diploïdes, et qu'un 4x peut produire 6 gamètes diploïdes, le tout uniformément. La Figure 3.4 ci-dessous donne à titre d'exemple les schémas de reproduction associés aux croisements de type 3x/3x, qui peuvent donc engendrer des enfants 2x, 3x ou même 4x : de 2 parents triploïdes, on peut voir qu'il en découle 36 enfants potentiels dont 9 sont diploïdes, 18 sont triploïdes et 9 sont tétraploïdes. (Le problème est essentiellement combinatoire !)

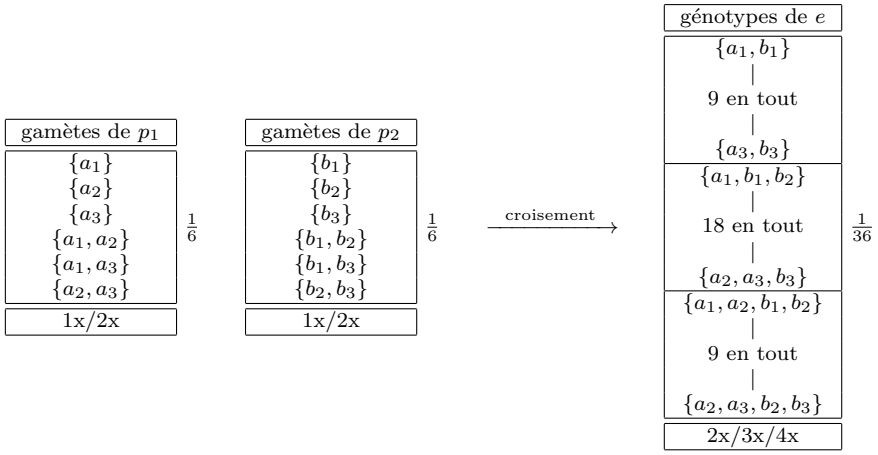


FIGURE 3.4 – Schémas de reproduction associés à une relation de la forme $(p_1, p_2) \mapsto e$ où les parents p_1 et p_2 sont triploïdes de génotypes $\{a_1, a_2, a_3\}$ et $\{b_1, b_2, b_3\}$, respectivement.

La seconde difficulté réside dans le fait que certaines données génotypiques ne sont pas connues avec certitude en présence de tri- ou tétraploïdie, en raison de la méthode de détection des allèles qui se fait par lecture de signaux en présence/absence. Un triploïde donnant lieu à un signal avec 2 pics en a et en b est soit $\{a, a, b\}$, soit $\{a, b, b\}$, et il en va de même lorsqu'un tétraploïde laisse apparaître 2 ou 3 pics. L'exemple donné sur la Figure 3.6 est issu du jeu de données, l'individu en question est tétraploïde et 2 pics sont détectés (132 et 161), ce qui signifie que son génotype à cet emplacement peut être aussi bien $\{132, 132, 132, 161\}$ que $\{132, 132, 161, 161\}$ ou que $\{132, 161, 161, 161\}$. Tenant compte de ces phénomènes, on évalue la probabilité d'une relation $(p_1, p_2) \mapsto e$ sur la base de m marqueurs par

$$\delta(e, p_1, p_2) = \prod_{s=1}^{m} \sum_{G \in \mathcal{G}_s} \mathbb{P}(\{(p_1, p_2) \mapsto e\} \,|\, G)\, \mathbb{P}(G) \tag{3.5}$$

où $\mathcal{G}_s$ est l'ensemble des génotypes possibles sur le signal s pour le triplet (e, p_1, p_2) et $\mathbb{P}(G)$ est la probabilité que l'on attribue à un génotype $G \in \mathcal{G}_s$ (qui reste source d'interrogations, comme on le verra dans les perspectives), alors que $\mathbb{P}(\{(p_1, p_2) \mapsto e\} \,|\, G)$ découle du modèle retenu pour les croisements (l'indépendance supposée entre les signaux est justifiée par une analyse statistique préalable dans laquelle on décorrèle les observations). Après renormalisation, on construit pour chaque individu une mesure de probabilité portée par l'ensemble des couples de $\mathcal{P}$ et quantifiant les liens de parenté directe potentiels. Cela permet en particulier de bâtir l'arbre généalogique le plus probable au sein de la population, à l'image de l'exemple proposé sur la Figure 3.5.

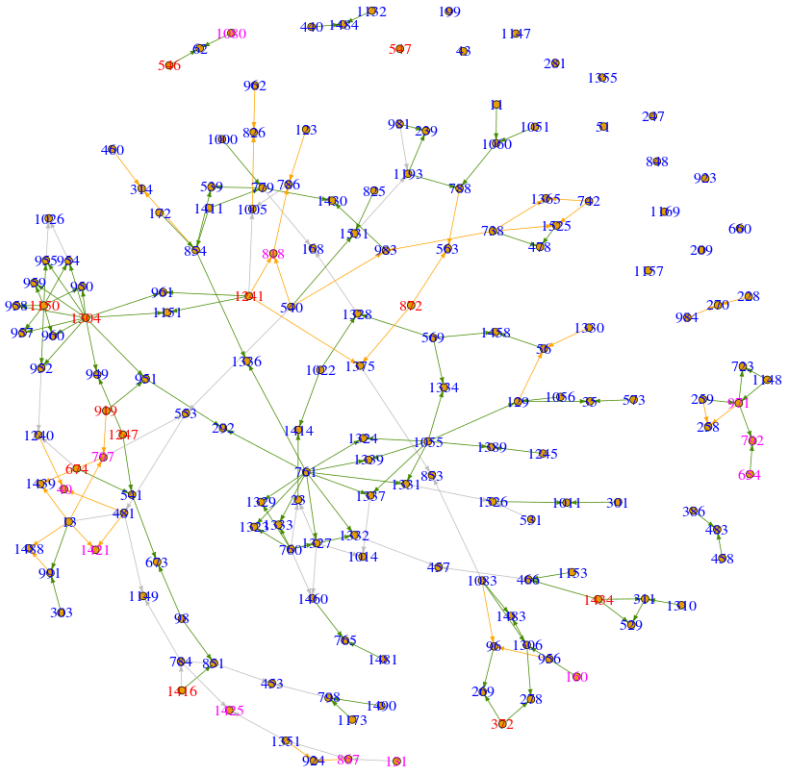


FIGURE 3.5 – Exemple de reconstruction de généalogies par maximum de vraisemblance. Les couleurs caractérisent la fiabilité des informations représentées.

Notre étude propose également une estimation de la loi de reproduction de chaque individu, c'est-à-dire du nombre d'enfants pour lesquels il est directement impliqué, en considérant l'information issue de toutes les généalogiques potentielles (et pas seulement celle de plus grande probabilité). Cette caractéristique est très importante du point de vue biologique car dans des populations entretenues, elle permet de comprendre rétrospectivement les variétés qui ont été favorisées comme géniteurs par les sélectionneurs en repérant les comportements atypiques. Enfin, un algorithme de recherche des chaînons manquants est mis en place afin de tenter de retrouver le génotype d'individus non observés mais vraisemblablement actifs par le passé. Nous sommes pour cela amenés à des considérations techniques justifiant la nécessité de définir un critère pénalisé pour décider si un individu apporte une information significative à la généalogie, à travers ce que l'on définit comme son potentiel d'interaction.

Perspectives

Le point d'amélioration qui ressort le plus souvent de ce travail, et qui par ailleurs est une piste de recherche active dans ce domaine particulier de la biologie végétale, porte sur le dosage allélique, c'est-à-dire sur l'évaluation de la probabilité $\mathbb{P}(G)$ dans (3.5). Il existe des arguments de nature purement technique pour affirmer que l'amplitude des pics sur les signaux n'est pas suffisante pour décider avec certitude qu'un allèle est plus présent qu'un autre. Repensons à l'exemple déjà évoqué et caractéristique de cette situation

ambiguë : un tétraploïde dont l'étude en présence/absence révèle la présence des allèles a et b pourrait assez naturellement se voir attribuer le génotype $\{a, b, b, b\}$ si le pic en b est trois fois plus grand que le pic en a, comme sur la Figure 3.6 ci-dessous issu des données réelles.

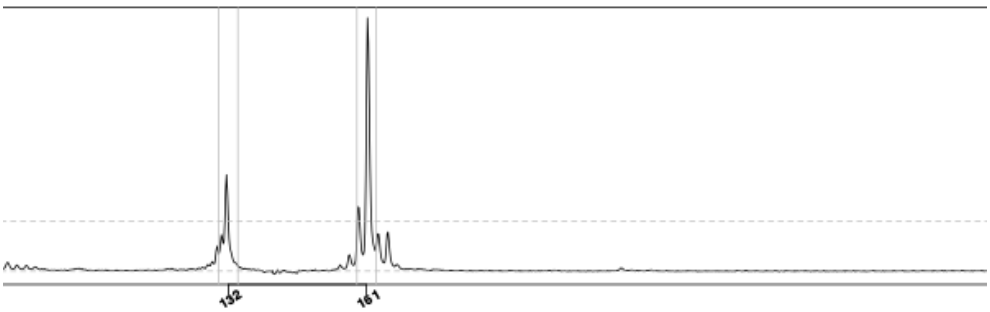


FIGURE 3.6 – Exemple de signal pour un marqueur microsatellite particulier donnant lieu à la détection de 2 pics (allèles 132 et 161) pour un individu tétraploïde.

Des études sont en cours, on propose en particulier de déterminer les probabilités de dosage par une agrégation de l'estimation obtenue de manière purement matérielle (par lecture des signaux) avec celle obtenue de manière purement génétique (rapport des allèles dans une population en équilibre). La piste est très prometteuse mais elle bute encore sur un obstacle de taille : l'absence d'échantillon de validation. Par ailleurs, nous avons retenu en première analyse le modèle le plus intuitif concernant la reproduction (uniformité généralisée) mais selon les experts biologistes, certaines incongruités peuvent survenir dans le monde végétal : l'appariement préférentiel, la double réduction, l'infertilité potentielle des triploïdes (bien que sélectionnés et favorisés... pour leur rareté !), etc. En bref, il pourrait être intéressant de développer des schémas de reproduction plus généraux, ce qui ne semble pas une tâche compliquée d'un point de vue conceptuel bien que nécessitant l'introduction de nouveaux paramètres d'ajustement. L'algorithme de recherche des chaînons manquants est de type 'glouton' (exploratoire) et en ce sens très coûteux en temps de calcul, là encore des pistes d'améliorations pourraient être creusées. Sous l'angle de vue probabiliste, une telle étude renvoie immédiatement à des problématiques plus ambitieuses comme les équilibres d'Hardy-Weinberg ou encore la coalescence. À ce sujet, on sait finalement assez peu de choses dans les populations tétraploïdes en dehors de l'étude de Arnold *et al.* (2012) et beaucoup de points restent à éclaircir. Pour conclure, permettons-nous de citer l'intervenant biologiste principal (J. Clotault) : sur un échantillon de rosiers 2x, 3x et 4x pourtant choisis pour être assez distants génétiquement et apparus tout au long du XIXe siècle, l'application de la méthode KING (Manichaikul *et al.* (2010)) de calcul d'apparentement, dans le sens évoqué plus haut, suggère de façon surprenante un apparentement fort entre la plupart de ces rosiers et il serait très intéressant de pouvoir valider (ou non...) les résultats de cette méthode avec une méthode plus fiable, basée sur les lois de l'hérédité, comme la nôtre. L'étude est en cours.

PROBABILISTIC RECONSTRUCTION OF GENEALOGIES FOR POLYPLOID PLANT SPECIES

PROÏA FRÉDÉRIC, PANLOUP FABIEN, TRABELSI CHIRAZ, AND CLOTAULT JÉRÉMY

Abstract. A probabilistic reconstruction of genealogies in a polyploid population (from 2x to 4x) is investigated, by considering genetic data analyzed as the probability of allele presence in a given genotype. Based on the likelihood of all possible crossbreeding patterns, our model enables us to infer and to quantify the whole potential genealogies in the population. We explain in particular how to deal with the uncertain allelic multiplicity that may occur with polyploids. Then we build an *ad hoc* penalized likelihood to compare genealogies and to decide whether a particular individual brings sufficient information to be included in the taken genealogy. This decision criterion enables us in a next part to suggest a greedy algorithm in order to explore missing links and to rebuild some connections in the genealogies, retrospectively. As a by-product, we also give a way to infer the individuals that may have been favored by breeders over the years. In the last part we highlight the results given by our model and our algorithm, firstly on a simulated population and then on a real population of rose bushes. Most of the methodology relies on the maximum likelihood principle and on graph theory.

1. Introduction

1.1. **Motivations.** Pedigrees depict the genealogical relationships between individuals of a given population. They can be built thanks to mating knowledge or they can be inferred from molecular markers. The identification of pedigrees allows a broad variety of applications: genealogy identification, like in grapevine [12], improvement of conservation programs for endangered species [14], inference of statistics used in quantitative and population genetics like heritability or population effective size [1, 10], etc. Like for most population genetics analyses, pedigree reconstruction methods and their implementation were firstly developed for diploid species (but see [21]). Polyploids, *i.e.* species with more than two alleles for a given locus, represent approximately 25% of plant species [2], and among them a large number of cultivated species. Polyploidy in animals is more rare but some examples were described in insects, fishes, amphibians and reptiles [18, 15].

Several strategies were used to reconstruct the genealogical relationships from molecular markers (reviewed in [9]). Exclusion methods eliminate potential parents which do not show at least one allele per locus shared with a putative offspring. If more than two parents are possible, categorical allocation methods allow identification of the most likely parents according to their probability to transmit alleles shared with the potential progeny. Parental reconstruction methods use full- or half-siblings in order to identify the most likely parents. By comparison, sibling reconstruction methods add a preliminary step of inference of siblings when they are unknown. In this paper, the objective is to adapt and to extend the approach

Key words and phrases. Allelic multiplicity, Crossbreeding patterns, Genealogies of plant species, Graph theory, Maximum likelihood principle, Missing links, Pedigree reconstruction, Polyploid population.

of [4], namely to determine for each individual the most likely couple of parents amongst all older individuals, so as to build some family trees in polyploid plant species. Our study certainly intends to be applied on real genetic datasets, in particular the main practical motivation is to find some retrospective links in a population of rose bushes that will now be described.

The empirical dataset used in the last section of this article was obtained on cultivated roses bred mainly during the nineteenth century (*Rosa* sp.). Rose breeding activities were particularly abundant during the nineteenth century and were very documented. As an example, breeding year is known for a majority of roses from this period. However, the genealogical relationships described in archives are highly hypothetical, due to the lack of control of artificial hybridization until the end of the nineteenth century. Among the approximately 200 species of the genus *Rosa*, ploidy level varies between 2x and 10x [8]. Rose breeding activities from the nineteenth century involved interspecific crossings between diploid species and tetraploid species, with a small contribution of genotypes with higher ploidy like species from the *Caninae* section (4x, 5x and 6x) [17, 13]. Cultivated roses bred during the nineteenth century can exhibit all ploidy values between 2x and 6x, even if 5x and 6x are rare [13]. The mode of inheritance in these rose cultivars remains highly unknown. It is generally considered that modern tetraploid cultivated roses exhibit a tetrasomic inheritance (no preferred pairing among the set of four homologous chromosomes and creation of tetravalents during meiosis) [11]. But a mixture of disomic (preferred pairing of two bivalent pairs during meiosis according to their genomic similarity) and tetrasomic inheritance could be observed according to chromosomes and according to genotypes [3]. Triploid roses have played a major role in rose hybridizations. Like in other species, triploid roses exhibit a low fertility rate, due to irregular meiosis leading to aneuploidy [16]. However, even if the production of fertile gametes from triploids remains rare, these events were selected by breeders, especially as bridges between different ploidy levels. For example, *Bourbon*, *Hybrid China* and *Hybrid Tea* rose groups were both obtained by a cross between a Chinese diploid cultivar and a European tetraploid cultivar. First cultivars from these groups were triploid [7]. Triploids form both haploid and diploid gametes [20]. Following the obtention of a variety by hybridization, it was then propagated vegetatively by cutting or grafting and often conserved in rose gardens. Therefore rose varieties can be considered as immortal and they could have been involved at different periods in rose pedigrees. As most of plants, roses are hermaphrodites and can therefore have been used as female or male on different hybridization events. Selfing rate in roses is very low mainly because of self-incompatibility ([19] and J. Mouchotte, pers. comm.). These specific breeding behaviors are the cornerstone of our probabilistic model.

In a general way, the polyploidy of the population may give rise to complications in terms of multiplicity of the alleles, being only aware of their presence or absence: that will be one of our strategic challenges to deal with this lack of information, widely discussed throughout the manuscript. Whereas for diploids the presence or absence of alleles is sufficient – for $\{a\}$ and $\{a, b\}$ undoubtedly correspond to $\{a, a\}$ and $\{a, b\}$ – the observation of $\{a, b\}$ for a tetraploid can correspond to $\{a, a, a, b\}$, $\{a, a, b, b\}$ or $\{a, b, b, b\}$. Reading the presence or absence of alleles on electrophoregrams and interpreting theoretical ratios between peak intensities is an option to determine the number of copies of each allele [6]. Unfortunately, we will explain in good time the reasons why this strategy is not reliable in our context and

we will introduce a way to deal with this allelic multiplicity through the intermediary of probabilities related to each configuration. Before getting to the heart of the matter, let us point out that the objective of this work is not to introduce a biological issue, but rather to build and justify the more realistic mathematical framework regarding the biological model of roses bred during the nineteenth century. This work is above all a methodological one.

The paper is organized as follows. In Section 2, we present a probabilistic method in order to reconstruct genealogies for species with several ploidy levels, from 2x to 4x, by considering genetic data analyzed as the probability of allele presence in a given genotype. In particular, we compute the likelihood associated with all crossbreeding patterns and we explain how to build and quantify the whole possible genealogies of the population and how to treat the unknown allelic multiplicity. As a by-product we also give a way to find the individuals favored by breeders, retrospectively. Section 3 treats the isolated individuals, more precisely, the missing links. Under some criteria, we suggest an algorithm computing virtual individuals to improve the genealogy. Whereas Sections 2 and 3 are mainly theoretical, all our results will be tested in Section 4, both on a simulated population and on a rose bushes population. We conclude by highlighting some weaknesses of our methodology and by giving, in accordance, some trails for future studies.

1.2. **Preliminary considerations and notations.** In the whole paper, $\mathcal{P}$ stands for the population of size $n = \mathrm{Card}(\mathcal{P})$ and m is the number of genes involved in the reconstruction process. Technically, m corresponds to the number of signals on which we read the *peaks*, expressing the set of alleles detected on each gene. We make the crucial hypothesis that signals are *mutually independent*, which can be argued on a genetic as well as statistical point of view (genes are chosen for their absence of known interaction and a prior statistical treatment tends to decorrelate them by eliminating material-type influences). For an individual $e \in \mathcal{P}$, we denote by $g_s(e)$ the *genotype* of gene s, that is, the *set of alleles* present for this gene, shortened in $g(e)$ when we deal with an unspecified gene (to be precise, we should in fact speak of *multiset* since we may have multiple instances of the same allele in the genotype, however we shall not make these kind of distinctions). We also denote by $x(e) = \mathrm{Card}(g(e)) \in \{2, 3, 4\}$ the *ploidy* of e, the number of sets of chromosomes in a cell. In addition, we assume that the birth dates are known and that no death occurs, which is consistent with the fact that the work is related to plant cultivars. We also assume that gametes are produced according to strict polysomic inheritance and we neglect double reduction.

2. Likelihood of a genealogy

This section is the heart of the paper. Firstly we will describe the genetic patterns that we retain to cross the polyploid individuals, and we will discuss the probabilistic treatment of the allelic multiplicity that may appear for triploids and tetraploids. Thereafter, we will be in the position to estimate some retrospective links and to compute an *ad hoc* penalized likelihood for the genealogy. Before anything else, let us begin with a formal description of what we mean by genealogy and likelihood. A genealogy is an element of the set

$$\Upsilon(\mathcal{P}) = \prod_{e \in \mathcal{P}} \left\{ \mathbb{T}(e) \cup (e, \varnothing) \right\} \quad \text{where} \quad \mathbb{T}(e) = \bigcup_{s \in \mathcal{S}(e)} (e, s) \tag{2.1}$$

and where $\mathcal{S}(e)$, as will be detailed in good time (see beginning of Subsection 2.3), is the set of non-ordered pairs candidates to the genealogy of e. In concrete terms, an individual

e is associated with each couple of possible parents $\mathcal{S}(e)$ together with $\varnothing$, to cover the case where $\mathbb{T}(e) = \varnothing$, that is where no triplet offspring/couple of parents can be found in the population for e. Thus, a genealogy $\mathcal{T}$ on $\mathcal{P} = \{e_1, \ldots, e_n\}$ takes the form of

$$\mathcal{T} = \big\{(e_1, s_1),\, (e_2, s_2),\, \ldots,\, (e_n, s_n)\big\} \tag{2.2}$$

in which s_i is either an element of $\mathcal{S}(e_i)$, either $\varnothing$. It has clearly a structure of graph, as will be explained later. Now, looking at $\mathcal{T}$ as the realization of a discrete random vector taking values in the set $\Upsilon(\mathcal{P})$, it naturally follows that the *likelihood of a genealogy* is the probability that it has to be observed, in accordance with the statistical usual definition, given a model and a set of hypotheses that will be described in this section. It should also be noted that a maximum likelihood genealogy, as will be largely discussed later, is not an estimator in the statistical sense, but the value of $\mathcal{T} \in \Upsilon(\mathcal{P})$ having the biggest probability, with respect to the model.

2.1. **Crossbreeding patterns.** To simplify the combinatorial analysis, we use the following natural models. Diploids produce haploid gametes, genotype $\{a, b\}$ leads to gametes $\{a\}$ and $\{b\}$ with probability 1. Triploids produce haploid and diploid gametes, genotype $\{a, b, c\}$ leads to gametes $\{a\}$ and $\{b, c\}$ with probability $\frac{1}{3}$, gametes $\{b\}$ and $\{a, c\}$ with probability $\frac{1}{3}$ and gametes $\{c\}$ and $\{a, b\}$ with probability $\frac{1}{3}$. Tetraploids produce diploid gametes, genotype $\{a, b, c, d\}$ leads to gametes $\{a, b\}$ and $\{c, d\}$ with probability $\frac{1}{3}$, gametes $\{a, c\}$ and $\{b, d\}$ with probability $\frac{1}{3}$ and gametes $\{a, d\}$ and $\{b, c\}$ with probability $\frac{1}{3}$. In addition, each individual can either be male or female, the set of gametes is treated as an *urn problem*. Crossing is made by choosing at random two gametes among all these possibilities, bringing them together to obtain the offspring's genotype. Figures 1–2–3 are schematic representations of the gametes production, indicated by arrows, of a parent cell.

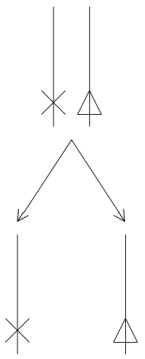


FIGURE 1. Schematic representation of the gametes production (in the bottom) for a diploid cell (in the top). Symbols represent the alleles of a given gene on its chromosome (line).

Let p_1 and p_2 be two individuals having ploidies $x(p_1)$ and $x(p_2)$ with genotypes $g(p_1) = \{a_1, \ldots, a_{x(p_1)}\}$ and $g(p_2) = \{b_1, \ldots, b_{x(p_2)}\}$, respectively. In the sequel, p_1 and p_2 are the parents of the offspring e. The different ploidy levels lead to six patterns that we are now going to describe in detail.

(P_1) $x(p_1) = x(p_2) = 2$. Let $g(p_1) = \{a_1, a_2\}$ and $g(p_2) = \{b_1, b_2\}$. Then, e has 4 potential diploid genotypes $g(e) = \{a_i, b_k\}$, for $i, k \in \{1, 2\}$. Each one has probability $\frac{1}{4}$.

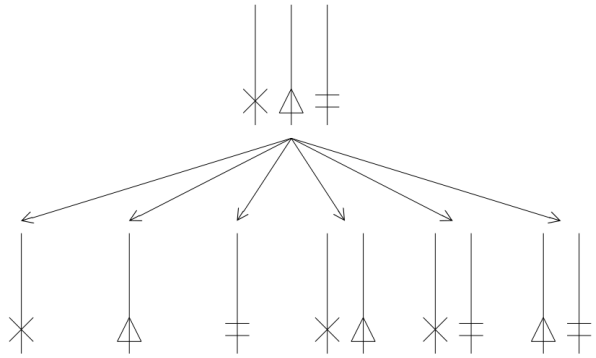


FIGURE 2. Schematic representation of the gametes production (in the bottom) for a triploid cell (in the top). Symbols represent the alleles of a given gene on its chromosome (line).

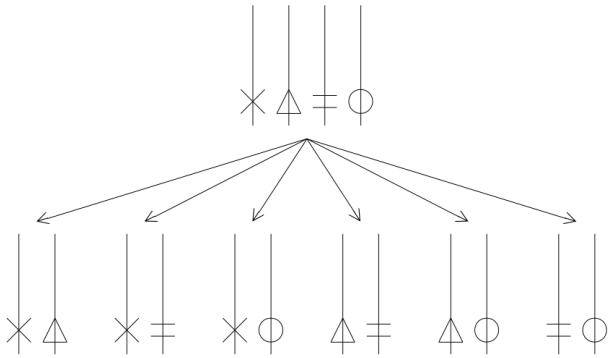


FIGURE 3. Schematic representation of the gametes production (in the bottom) for a tetraploid cell (in the top). Symbols represent the alleles of a given gene on its chromosome (line).

(P_2) $x(p_1) = 2$ and $x(p_2) = 3$. Let $g(p_1) = \{a_1, a_2\}$ and $g(p_2) = \{b_1, b_2, b_3\}$. Then, e has 6 potential diploid genotypes $g(e) = \{a_i, b_k\}$, and 6 potential triploid genotypes $g(e) = \{a_i, b_k, b_\ell\}$, for $i \in \{1, 2\}$ and $k, \ell \in \{1, 2, 3\}$. Each one has probability $\frac{1}{12}$.

(P_3) $x(p_1) = 2$ and $x(p_2) = 4$. Let $g(p_1) = \{a_1, a_2\}$ and $g(p_2) = \{b_1, b_2, b_3, b_4\}$. Then, e has 12 potential triploid genotypes $g(e) = \{a_i, b_k, b_\ell\}$, for $i \in \{1, 2\}$ and $k, \ell \in \{1, 2, 3, 4\}$. Each one has probability $\frac{1}{12}$.

(P_4) $x(p_1) = x(p_2) = 3$. Let $g(p_1) = \{a_1, a_2, a_3\}$ and $g(p_2) = \{b_1, b_2, b_3\}$. Then, e has 9 potential diploid genotypes $g(e) = \{a_i, b_k\}$, 18 potential triploid genotypes $g(e) = \{a_i, b_k, b_\ell\}$ or $g(e) = \{a_i, a_j, b_k\}$, and 9 potential tetraploid genotypes $g(e) = \{a_i, a_j, b_k, b_\ell\}$, for $i, j, k, \ell \in \{1, 2, 3\}$. Each one has probability $\frac{1}{36}$.

(P_5) $x(p_1) = 3$ and $x(p_2) = 4$. Let $g(p_1) = \{a_1, a_2, a_3\}$ and $g(p_2) = \{b_1, b_2, b_3, b_4\}$. Then, e has 18 potential triploid genotypes $g(e) = \{a_i, b_k, b_\ell\}$, and 18 potential tetraploid genotypes $g(e) = \{a_i, a_j, b_k, b_\ell\}$, for $i, j \in \{1, 2, 3\}$ and $k, \ell \in \{1, 2, 3, 4\}$. Each one has probability $\frac{1}{36}$.

(P$_6$) $x(p_1) = x(p_2) = 4$. Let $g(p_1) = \{a_1, a_2, a_3, a_4\}$ and $g(p_2) = \{b_1, b_2, b_3, b_4\}$. Then, e has 36 potential tetraploid genotypes $g(e) = \{a_i, a_j, b_k, b_\ell\}$, for $i, j, k, \ell \in \{1, 2, 3, 4\}$. Each one has probability $\frac{1}{36}$.

To sum up, all diploid offsprings may come from patterns (P$_1$)–(P$_2$)–(P$_4$), all triploid offsprings from patterns (P$_2$)–(P$_3$)–(P$_4$)–(P$_5$) and all tetraploid offsprings from patterns (P$_4$)–(P$_5$)–(P$_6$). One can remark that the trickiest case is probably (P$_4$) since three different ploidies can be generated by crossing triploids, Figure 4 gives a streamlined representation of it. Now let $\{(p_1, p_2) \mapsto e\}$ be the event through which the pair (p_1, p_2) conceives e, let u

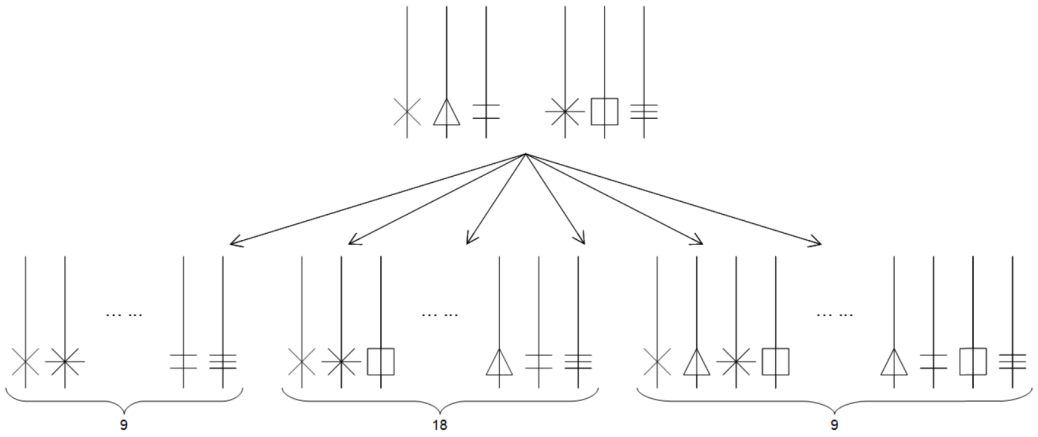


FIGURE 4. Schematic representation of pattern (P$_4$) leading to $u = 36$ potential offsprings including 9 diploids, 18 triploids and 9 tetraploids. Symbols represent the alleles of a given gene.

denote the maximum number of different genotypes generated by the pattern ($u = 4$, $u = 12$ or $u = 36$) corresponding to the ploidy of p_1 and p_2, and let $e_1, \ldots, e_u$ name the potential offsprings of the cross. Our hypotheses show that, conditionally on the knowledge of the genotypes of the parents, each offspring is drawn through a uniform distribution. So, we set

$$\mathbb{P}(\{(p_1, p_2) \mapsto e\} \,|\, \{g(p_1), g(p_2), g(e)\}) = \frac{1}{u} \sum_{r=1}^{u} \mathbb{1}_{\{e_r = e\}} \tag{2.3}$$

where the genetic equality $e_r = e$ means that $g(e_r)$ and $g(e)$ coincide in a sense that we have to define. Specifically, we consider that $e_r = e$ once

$$g(e_r) = g(e) \quad \text{and hence} \quad x(e_r) = x(e) \tag{2.4}$$

which in this case amounts to say that e_r and e have the same ploidy and the same set of alleles (we remind that $x = \text{Card}(g)$). However, it is important to highlight that (2.4) is only relevant from theoretical perspectives or on simulated data. We will see in Section 4.2 that real genotypes result from a calibration of the equipment and some rounded values to be interpreted as *base pairs*. Therefore,

$$x(e_r) = x(e) \quad \text{and} \quad \|g^*(e_r) - g^*(e)\|_\infty \leqslant 1 \tag{2.5}$$

where g^* stands for an ascending sorted vector containing the elements of g, should be an appropriate comparison on such data. Indeed, this criterion allows an offset of ± 1 base pairs for two corresponding alleles.

Examples. To illustrate this calculation method, let us consider $g(p_1) = \{a, a\}$ and $g(p_2) = \{a, b\}$. Then $u = 4$, the potential offsprings have genotypes $g(e_1) = g(e_3) = \{a, a\}$ and $g(e_2) = g(e_4) = \{a, b\}$. For $g(e) = \{a, a\}$ or $g(e) = \{a, b\}$, formula (2.3) gives probability $\frac{1}{2}$. It also gives probability 0 for all other genotypes. In the more intricate case where $g(p_1) = \{a, a, b, c\}$ and $g(p_2) = \{a, c, c\}$, then $u = 36$ and among the potential offsprings, 5 will have genotype $g(e) = \{a, b, c\}$. Formula (2.3) gives probability $\frac{5}{36}$ for such a triploid offspring.

2.2. **Allelic multiplicity.** For an individual $e \in \mathcal{P}$, the set $g(e)$ is the *true* genotype. However in our experimental studies, we only observe a *partial* genotype $\widehat{g}(e) \subset g(e)$ containing the distinct alleles – a set of peaks on the signal. Taking advantage of the ploidy $x(e)$, one is able to infer all possible $g(e)$ from $\widehat{g}(e)$. Explicitly, we use the following connections, where π names a probability of multiplicity in a generic way.

(C_1) $\widehat{g}(e) = \{a\}$ and $x(e) = 2$ leads to $g(e) = \{a, a\}$ with probability 1.
(C_2) $\widehat{g}(e) = \{a, b\}$ and $x(e) = 2$ leads to $g(e) = \{a, b\}$ with probability 1.
(C_3) $\widehat{g}(e) = \{a\}$ and $x(e) = 3$ leads to $g(e) = \{a, a, a\}$ with probability 1.
(C_4) $\widehat{g}(e) = \{a, b\}$ and $x(e) = 3$ leads to $g(e) = \{a, a, b\}$ with probability π_{21} and to $g(e) = \{a, b, b\}$ with probability π_{12}. We set $\pi_{21} + \pi_{12} = 1$.
(C_5) $\widehat{g}(e) = \{a, b, c\}$ and $x(e) = 3$ leads to $g(e) = \{a, b, c\}$ with probability 1.
(C_6) $\widehat{g}(e) = \{a\}$ and $x(e) = 4$ leads to $g(e) = \{a, a, a, a\}$ with probability 1.
(C_7) $\widehat{g}(e) = \{a, b\}$ and $x(e) = 4$ leads to $g(e) = \{a, a, a, b\}$ with probability π_{31}, $g(e) = \{a, a, b, b\}$ with probability π_{22} and $g(e) = \{a, b, b, b\}$ with probability π_{13}. We set $\pi_{31} + \pi_{22} + \pi_{13} = 1$.
(C_8) $\widehat{g}(e) = \{a, b, c\}$ and $x(e) = 4$ leads to $g(e) = \{a, a, b, c\}$ with probability π_{211}, $g(e) = \{a, b, b, c\}$ with probability π_{121} and $g(e) = \{a, b, c, c\}$ with probability π_{112}. We set $\pi_{211} + \pi_{121} + \pi_{112} = 1$.
(C_9) $\widehat{g}(e) = \{a, b, c, d\}$ and $x(e) = 4$ leads to $g(e) = \{a, b, c, d\}$ with probability 1.

Instead of selecting a genotype for e when several are conceivable, that is, for combinations (C_4)–(C_7)–(C_8), the model that we introduce in the next section takes account of all possibilities weighted by their related probabilities. In fact, our model enables us to choose if necessary $\pi = \pi^{(s)}$ gene by gene or, equivalently, signal by signal, to consider the different interpretations of the relative amplitude of the peaks on each signal, for material reasons. We will describe it in more details in the beginning of Section 4.2.

2.3. **Probability of a genealogical link.** For any individual $e \in \mathcal{P}$, as it has been outlined in the introduction of the section, let $\mathcal{S}(e) \subset \mathcal{P}^2$ be the *compatible* subpopulation, that is, the set of non-ordered pairs (p_1, p_2) with $p_1 \neq p_2$ (excluding selfing) genetically and chronologically candidates to the genealogy of e. It is worth noting that the only chronological constraint is obviously to consider that birth dates of descendants cannot be prior to the ones of their parents. In particular, the probabilities of ancestry are considered as *time-invariant*: any individual has the same probability of being a parent, regardless of its birth date, excluding *de facto* any generational model like Galton-Watson trees. This point of view is specific to plant species, and would clearly be irrelevant for animal populations. Whether

the individual was obtained during the decade preceding the birth date of the offspring, or several centuries ago, because of the immortality and constant fertility given by a vegetative propagation, we assume that the probability of ancestry is the same. Our objective is to build a probability measure on $\mathcal{S}(e) \cup \{\varnothing\}$ quantifying the whole possible genealogical links of e, the element $\varnothing$ being added to cover the case where no parents can be found in the population. The hypothesis of mutual independence of the signals allows us to work on each signal and to multiply the results. Let

$$\delta(e, p_1, p_2) = \prod_{s=1}^{m} \sum_{G \in \mathcal{G}_s} \mathbb{P}(\{(p_1, p_2) \mapsto e\} \,|\, G)\, \mathbb{P}(G) \tag{2.6}$$

where $\mathcal{G}_s$ is the set of all possible genotypes on signal s for the triplet (e, p_1, p_2). In the best-case scenario, $\mathrm{Card}(\mathcal{G}_s) = 1$ which means that $\widehat{g}_s(p_1)$, $\widehat{g}_s(p_2)$ and $\widehat{g}_s(e)$ lead to no uncertain allelic multiplicity, and thus $\mathbb{P}(G) = 1$. At worst, $\mathrm{Card}(\mathcal{G}_s) = 27$ meaning that $\widehat{g}_s(p_1)$, $\widehat{g}_s(p_2)$ and $\widehat{g}_s(e)$ are in the situation (C_7) or (C_8), and $\mathbb{P}(G)$ is the product of the related probabilities.

Example. Suppose that $x(p_1) = 3$, $x(p_2) = 4$, $x(e) = 4$ and that, on a particular signal s, we observe $\widehat{g}_s(p_1) = \{a, b\}$, $\widehat{g}_s(p_2) = \{a, c, d\}$ and $\widehat{g}_s(e) = \{a, d\}$. Then, $\mathrm{Card}(\mathcal{G}_s) = 18$. Indeed, we build $\mathcal{G}_s$ by combining $\{a, a, b\}$ and $\{a, b, b\}$ for p_1, $\{a, a, c, d\}$, $\{a, c, c, d\}$ and $\{a, c, d, d\}$ for p_2, and $\{a, a, a, d\}$, $\{a, a, d, d\}$ and $\{a, d, d, d\}$ for e. For the first combination we have $\mathbb{P}(G) = \pi_{21}^{(s)}\, \pi_{211}^{(s)}\, \pi_{31}^{(s)}$, for the second one $\mathbb{P}(G) = \pi_{21}^{(s)}\, \pi_{211}^{(s)}\, \pi_{22}^{(s)}$, and so on.

It only remains to renormalize. Explicitly, with

$$\Delta(e) = \sum_{(p_1, p_2) \in \mathcal{S}(e)} \delta(e, p_1, p_2) \tag{2.7}$$

where $\delta(e, p_1, p_2)$ is given in (2.6), let

$$\forall\, (p_1, p_2) \in \mathcal{S}(e), \quad \nu_e((p_1, p_2)) = \begin{cases} \frac{\delta(e, p_1, p_2)}{\Delta(e)} & \text{if } \Delta(e) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

and fix $\nu_e(\varnothing) = 1$ as soon as $\Delta(e) = 0$, and $\nu_e(\varnothing) = 0$ otherwise. Then clearly, $\nu_e : \mathcal{S}(e) \cup \{\varnothing\} \to [0, 1]$ is a probability measure that can be applied to look for the whole genealogy of $e \in \mathcal{P}$. To build the *most likely genealogy*, we must pick

$$c^*(e) = \underset{c \in \mathcal{S}(e) \cup \{\varnothing\}}{\arg\max}\ \nu_e(c). \tag{2.9}$$

To be precise, $c^*(e)$ defined as above is not necessarily unique, in such case we arbitrarily pick one optimum at random. We will see in the sequel that choosing a genealogical link amongst others is not necessarily relevant, hence we also consider

$$G(e) = \{c \in \mathcal{S}(e) \cup \{\varnothing\} \mid \nu_e(c) > 0\} \tag{2.10}$$

which represents the whole potential genealogical links of e in our population $\mathcal{P}$.

2.4. **A retrospective family tree.** Now the objective is to compute $G(e)$ – and thus $c^*(e)$ – for all $e \in \mathcal{P}$. In the framework of this study, a *family tree* $\mathcal{T}$ of the population $\mathcal{P}$ is a set of triplets (e, p_1, p_2) having probabilities $\nu_e((p_1, p_2)) > 0$, on the basis of m genes, such that there is at most one triplet (e, p_1, p_2) for any individual e, interpretable as the realization of the event $\{(p_1, p_2) \mapsto e\}$, taking up the notation of the previous sections. We also require

that a triplet (e, p_1, p_2) is assigned to the node e of the family tree as soon as $c^*(e) \neq \varnothing$, that is as soon as there exists at least one potential genealogical link for e. To make the connection with our formal introduction, a family tree $\mathcal{T}$ completed by $(e, \varnothing)$ for each e such that $c^*(e) = \varnothing$ is merely a genealogy as it is defined in (2.2). In an equivalent way, we build a *graph* in which each individual is a vertex and each genealogical link is a couple of arcs (from the parents to the offspring). Note that the chronological constraint applied on $\mathcal{S}(e)$ is sufficient to ensure that no cycle is present in the graph. The methods and algorithms that follow will be tested and applied in Section 4.

2.4.1. *Most likely trees.* Combining all options of $G(e)$ for each $e \in \mathcal{P}$ gives an exhaustive set of trees, all potential genealogies of the population that we will denote as $\mathbb{G}(\mathcal{P})$ in (2.13). However, on large datasets, this can be difficult due to the exponential growth of the combinations. Thus we look for criteria of selection, and first we define the log-likelihood of a family tree $\mathcal{T}$ as follows,

$$\ell(\mathcal{T}) = \sum_{(e,p_1,p_2) \in \mathcal{T}} \ln \nu_e((p_1, p_2)). \tag{2.11}$$

Note that this expression corresponds to the likelihood of a genealogy as we have defined it beforehand, under the crucial hypothesis that each triplet offspring/couple of parents is independent of any other, which once again is specific to plant species. Clearly $\mathcal{P}$ can be divided into $\mathcal{L} = \{e \in \mathcal{P} \,|\, c^*(e) \neq \varnothing\}$ and $\mathcal{I} = \{e \in \mathcal{P} \,|\, c^*(e) = \varnothing\}$, respectively the individuals having potential ancestors in the population, present as nodes in all family trees built according to our constraints, and the ones for which we have not been able to find any genealogical link, that we will describe as *isolated*. Our model guarantees that maximizing $\ell(\mathcal{T})$ amounts to locally maximizing the log-probability of each link. To sum up,

$$\max_{\mathcal{T} \in \Upsilon(\mathcal{P})} \ell(\mathcal{T}) = \sum_{e \in \mathcal{L}} \ln \nu_e(c^*(e)) \tag{2.12}$$

and this upper bound is reached by the tree $\mathcal{T}^*$ built on all $e \in \mathcal{L}$ associated with the pairs $c^*(e)$. We shall note that formula (2.12) does not necessarily highlight a unique family tree, for some pairs (p_1, p_2) may have the same probability of producing e. In this case, the maximization problem has more than one solution.

2.4.2. *Number of offsprings.* Suppose now that the population is small enough to be able to compute

$$\mathbb{G}(\mathcal{P}) = \prod_{e \in \mathcal{P}} G(e) \tag{2.13}$$

where $G(e)$ is given in (2.10). Namely, $\mathbb{G}(\mathcal{P})$ contains the exhaustive set of potential genealogies of the population. Due to the combination of the options of all $G(e)$, $\mathrm{Card}(\mathbb{G}(\mathcal{P}))$ may be very large. In fact such a Cartesian product is only conceptual, but quickly intractable for practical purposes leading to combinatorial explosions. Therefore, a *threshold probability* must be used to select the genealogies of $\mathbb{G}(\mathcal{P})$. Concretely, we can replace the definition of $G(e)$ in (2.10) by the more stringent

$$G(e) = \{c \in \mathcal{S}(e) \cup \{\varnothing\} \mid \nu_e(c) > \pi_{\min}\} \tag{2.14}$$

for a given choice of $0 \leqslant \pi_{\min} < 1$, and the construction of $\mathbb{G}(\mathcal{P})$ accordingly. If we define $N(i)$ as a random variable counting the offsprings of $i \in \mathcal{P}$, then it could be interesting to

give an estimation of its probability distribution so as to infer, retrospectively, the individuals favored by breeders. Our model directly suggests to use

$$\forall\, k \in \mathbb{N}, \quad \widehat{\mathbb{P}}(N(i) = k) = \sum_{g \,\in\, \mathbb{G}(\mathcal{P})} w_g \, \mathbb{1}_{\{n_g(i) = k\}} \tag{2.15}$$

where $n_g(i)$ is the number of offsprings of i in the genealogy g and w_g is a *weighting* of the genealogy that can naturally be defined as the ratio between its likelihood and the sum of all likelihoods, *i.e.*

$$w_g = \frac{\mathrm{e}^{\,\ell(g)}}{L(\mathcal{P})} \quad \text{with} \quad L(\mathcal{P}) = \sum_{h \,\in\, \mathbb{G}(\mathcal{P})} \mathrm{e}^{\,\ell(h)} \tag{2.16}$$

keeping the notation of (2.11). It follows that

$$\widehat{\mathbb{E}}[N(i)] = \sum_{g \,\in\, \mathbb{G}(\mathcal{P})} w_g \, n_g(i) \tag{2.17}$$

may be a useful tool to decide whether i has been favored by breeders, by comparison with the global mean value and a classical *outlier threshold*. This approach will be illustrated on the rose bushes population of Section 4.2.

Example. Consider a set of 4 genealogies of likelihood 0.8, 0.6, 0.1 and 0.02, among which an individual i has 0, 1, 1 and 2 offsprings, respectively. Then we propose estimating $\widehat{\mathbb{P}}(N(i) = 0) \approx 0.526$, $\widehat{\mathbb{P}}(N(i) = 1) \approx 0.461$, $\widehat{\mathbb{P}}(N(i) = 2) \approx 0.013$ and $\widehat{\mathbb{P}}(N(i) > 2) = 0$. For this individual, $\widehat{\mathbb{E}}[N(i)] \approx 0.487$.

To look at pairwise relationships in the population, it can also be meaningful to build a *genealogical graph* made of all possible (weighted) links. In such a graph, we are not interested in the triplets offspring/couple of parents, but only in the pairs offspring/parent. For all $(i, j) \in \mathcal{P}^2$ and the same weights as in (2.16), consider

$$W_{i \to j} = \sum_{g \,\in\, \mathbb{G}(\mathcal{P})} w_g \, \mathbb{1}_{\{(i \to j) \,\in\, g\}} \tag{2.18}$$

where $\{(i \to j) \in g\}$ means that i is a parent of j in the genealogy g. The directed and weighted graph built on $W_{i \to j}$ amounts to the superposition of all genealogies except that the viewpoint is different: edges are not considered in pairs, but each one has a role of its own. However it is worth noting that, according to this model, the outflow from an individual is precisely its averaged number of offsprings (2.17). Thus, these two approaches are numerically equivalent but they differ from the interpretation.

2.4.3. *Comparison of trees.* For a fixed population of size n, since each tree contains the same number of links, maximizing the likelihood *via* (2.11) seems a suitable criterion. However, it cannot be trusted to compare trees with a different number of links. To understand this, let $\mathcal{P}_i = \mathcal{P} \cup \{i\}$ be the same population enhanced with a new individual, from the last generation, such that $\delta(i, p_1, p_2) > 0$ for at least two pairs $(p_1, p_2) \in \mathcal{S}(i)$. Then, for these pairs we get $\ln \nu_i((p_1, p_2)) < 0$, implying that $\ell(\mathcal{T}) > \ell(\mathcal{T}_i)$, where $\mathcal{T}$ and $\mathcal{T}_i$ are the family trees maximizing the likelihood on $\mathcal{P}$ and $\mathcal{P}_i$, respectively. In other words, this criterion favors $\mathcal{T}$ rather than $\mathcal{T}_i$ whereas there exists a link between some individuals of $\mathcal{P}$ and i. In

order to overcome this negative impact, as soon as we have to compare family trees on two populations $\mathcal{P}$ and $\mathcal{P}_i$ such that $\mathcal{P}_i = \mathcal{P} \cup \{i\}$, we suggest to consider a trade-off like

$$\ell^*(\mathcal{T}_i) = \ell(\mathcal{T}_i) + \Psi(i) \tag{2.19}$$

where $\ell(\mathcal{T}_i)$ is the log-likelihood given by (2.11) of the genealogical tree $\mathcal{T}_i$ on $\mathcal{P}_i$ containing i, and $\Psi(i)$ is a measure of the *interaction ability* of the new individual i with $\mathcal{P}$. Whence, to decide whether i has to be added into the genealogy, it will be possible to compare $\ell^*(\mathcal{T}_i)$ and $\ell(\mathcal{T})$ for the most likely tree $\mathcal{T}$ built on $\mathcal{P}$, provided a suitable adjustment of $\Psi(i)$. In this way, we intend to compensate the mechanical decrease of the log-likelihood due to the accumulation of potential links including i. This penalization of the log-likelihood is a strategy similar to the well-known AIC and BIC criteria. In the next section, when looking for missing individuals that could improve the family tree, we will see how to give a suitable explicit form to Ψ according to our purposes.

3. Missing links

Recall that our model assumes that no death occurs, which, as we have seen, is consistent with the fact that the work is related to perennial plant cultivars with asexual multiplication. However, individuals are obviously missing in the population – because they represent intermediate individuals never recorded as a cultivar and never distributed by the breeder, because the cultivar disappeared from rose gardens deliberately or accidentally, or because it was not sampled in the study. In this section, our objective is to look for some *missing links*. Since we do not know exactly how many individuals are missing, our strategy is to launch a *greedy algorithm* that explores the population and tries to detect an excess of information that might improve substantially the genealogy. The combinatorial complexity leads us to focus on some particular areas for the algorithm. More precisely, it seems that the isolated individuals are suitable starting points, for which we recall that $\mathcal{I} = \{e \in \mathcal{P} \,|\, c^*(e) = \varnothing\}$ is the set of individuals having no parents in the most likely genealogy. For all $e \in \mathcal{I}$, let $\mathcal{R}(e) \subset \mathcal{P}$ be the individuals in the population chronologically candidates to the genealogy of e and able to produce a gamete compatible with e. In addition, for each $p \in \mathcal{R}(e)$, consider

$$i^*(e,p) = \arg\max_i \, \delta(e,p,i) \tag{3.1}$$

as it is defined in (2.6), where i has the structure of an individual of the population (with a ploidy, a date of birth and a set of alleles for each signal). Namely, $i^*(e,p)$ is a virtual individual having a genotype which maximizes the probability of the event $\{(p,i) \mapsto e\}$, it can be seen as the "perfect partner" of p to produce e. Given $i = i^*(e,p)$, we now have to decide whether i significantly improves the genealogy. Let us carry on with the criterion introduced in (2.19), where the enhanced population is $\mathcal{P}_i = \mathcal{P} \cup \{i\}$. To match with our study, the penalization $\Psi(i)$ must favor individuals i providing the maximum number of interactions with $\mathcal{P}$. As we have seen in the last section, few interactions leave the likelihood almost unchanged whereas too many interactions tend to depreciate it, this was our motivation to look for a trade-off. We also want to give priority to any individual i reducing the number of *connected components* in the genealogy – that is, the number of subgraphs in which all nodes are connected. Indeed, in view of our fundamental hypothesis that, except for ancestors, all parents should be present in an ideal population, we know that if we were able to access to the whole population, it would lead to a graph with few connected components (less

than the number of ancestors, in any case). In this context, it seems natural to favor the reduction of the number of connected components, in order to get closer of this true (but inaccessible) genealogy. Define $\mathcal{T}$ and $\mathcal{T}_i$ as the maximum likelihood trees on $\mathcal{P}$ and $\mathcal{P}_i$, respectively, and suppose that i is contained in $\mathcal{T}_i$. Combining these requirements, we can write the penalization in the form

$$\Psi(i) = \lambda_i \frac{r(i)}{n} - \mu_i \Delta C(i) \tag{3.2}$$

where $r(i)$ is the number of individuals of $\mathcal{P}$ potentially interacting with i, $\Delta C(i)$ is the difference between the number of connected components in $\mathcal{T}$ and $\mathcal{T}_i$, $\lambda_i \geqslant 0$ and $\mu_i \geqslant 0$ are regularization parameters. Our decision rule consists in keeping an individual i which satisfies $\ell^*(\mathcal{T}_i) > \ell(\mathcal{T})$. We can formalize $r(i)$ like

$$r(i) = \sum_{p \in \mathcal{P}} \eta(i,p)$$

where $\eta(i,p) = 1$ if one can find $a \in \mathcal{P}$ such that $\delta(i,a,p) > 0$, $\delta(a,i,p) > 0$ or $\delta(p,a,i) > 0$, that is, if there is a nonzero probability for at least a link involving p and i, and $\eta(i,p) = 0$ otherwise. Note that a may be an offspring of i as well as a parent or a partner of i to be considered as an interaction involving i. To adapt our criterion, we can choose

$$\lambda_i = \frac{n}{2} \left| \ell(\mathcal{T}) - \ell(\mathcal{T}_i) \right| \tag{3.3}$$

since this guarantees that $\ell^*(\mathcal{T}_i) = \ell(\mathcal{T})$ when the new individual does not bring any connection except the one for which it has been created, not gathering connected components ($r(i) = 2$ and $\Delta C(i) = 0$), and thus when i should be rejected. A similar strategy enables us to fix μ_i, for $r(i) = 2$ must at least coincide with $\Delta C(i) = -1$ to make an interesting link. This is the case when i has been created to fulfill the event $\{(p,i) \mapsto e\}$, and when p and e belong to different connected components. Of course that situation must be favored, and to simplify one can choose

$$\mu_i = \lambda_i + 1 \tag{3.4}$$

which amounts to say that $\ell^*(\mathcal{T}_i) > \ell(\mathcal{T})$ whenever $\Delta C(i) < 0$. To enhance the population, we suggest the following algorithm.

(0) Fix $n_v > 0$, the maximum number of virtual individuals allowed to be inserted in the population.
(1) Build $\mathcal{R}(e)$ for all $e \in \mathcal{I}$.
(2) For all $p \in \mathcal{R}(e)$, compute the maximum likelihood partner i such that $\{(p,i) \mapsto e\}$ is achieved.
(3) Among these candidates, add in $\mathcal{P}$ the individual maximizing $\ell^*(\mathcal{T}_i)$ provided

$$\max_i \ \ell^*(\mathcal{T}_i) > \ell(\mathcal{T}).$$

Set $t_e - 1$ as birth date of the new individual, where t_e is the one of e.
(4) Recalculate the most likely tree $\mathcal{T}$ and the set $\mathcal{I}$ according to the new population.
(5) Repeat steps (1)–(4) as long as the criterion increases and $\mathrm{Card}(\mathcal{P}) < n + n_v$.

Before going further, let us focus on the complexity of this algorithm (and on some possible improvements). In the present state, it is fully exploratory and starts from arbitrary points. In terms of complexity, it is possible to evaluate that step (4) has a number of crosses

in the range of $O(n(n-1)(n-2))$ to be tested. Generally $\mathrm{Card}(\mathcal{I})$ is small, thus, even if it entirely depends on the population, let us suppose that it is bounded by $n_i \ll n$. The construction of $\mathcal{R}(e)$ requires $O(n-1)$ crosses to be tested for a given e. On the whole, we can roughly estimate that, considering a crossbreeding as the unit of measurement, $O(n_v\, n_i\, n(n-1)^2(n-2))$ operations are needed. In practice, much less operations are actually done since the symmetry and the chronological and genetical constraints cut a lot of paths. To reach a lower complexity, it should be relevant to look at less exploratory methods, in order to deal with the increasing number of individuals. In addition, the maximum of likelihood in step (2) is the natural solution, but it can also have unwelcome effects. In particular, this algorithm can not generate any triploid. This follows from the fact that, whenever a triploid produces a gamete, there exists a diploid or a tetraploid that produces the same gamete with a probability two times bigger. In the same vein, the virtual tetraploids can either be homozygous or heterozygous with only two distinct alleles. As a consequence, since the individual is specifically created to fulfill a particular crossbreeding, the situations where the missing link is a parent of more than one offspring in the population can not be recovered, except if the offsprings are genetically similar. This could be improved by testing not only the candidates, but also the mixes between them. For example, if $\{a, a, b, b\}$ is added to explain the presence of a diploid $\{a, b\}$, and if $\{c, c, d, d\}$ is added to explain the presence of another diploid $\{c, d\}$, then it could be interesting to add $\{a, b, c, d\}$ to explain both of them, instead. To conclude, we would like to highlight a last enhancement. Setting $t_e - 1$ as date of birth of the new individual is an arbitrary choice because, focusing on the offspring, we do not have any more information about the other interactions within the new genealogy. Each birth date between some initial time t_0 and $t_e - 1$ should be tested as well. All these improvements are hardly conceivable due to the computational complexity, except for small populations ($n \approx 50$, as in our simulations). Hence, as we can see, there are still numerous open questions to explore on the fundamental issue of the missing links.

4. An empirical study

The numerical processings were carried out through the `R` programming language and its software environment. In particular, we used the package *igraph*[1] to display the graphs. In all figures of this section, the geometric shapes that we use are circles to represent diploids, triangles for triploids and squares for tetraploids, gray individuals are real whereas white individuals are virtual. Similarly, we use solid lines for true links as well as dotted lines for the wrong links given by the model (unless noted otherwise). The computations are conducted *via* the uniform probabilities $\pi_{21} = \pi_{12} = \frac{1}{2}$ and $\pi_{31} = \pi_{22} = \pi_{13} = \pi_{211} = \pi_{121} = \pi_{112} = \frac{1}{3}$. The estimation of the mean number of offsprings is given by (2.17) and the *outlier threshold* is chosen to the standard $q_3 + 1.5\,(q_3 - q_1)$ with q_1 and q_3 the first and third quartiles of a subset of observations. It is computed using a moving window on the values in chronological order and then extrapolated by a linear regression, to take into account the time-invariance in the reproduction law and, thus, the fact that the older an individual is, the more offsprings he is likely to have.

4.1. **On a simulated population.** Consider the simulated population $\mathcal{P}$ whose detailed description is provided in the Appendix. To sum up, there are $n = 54$ individuals among

[1] `https://cran.r-project.org/web/packages/igraph/igraph.pdf`

which 17 diploids, 17 triploids and 20 tetraploids have interacted throughout 8 generations. The simulation relies on $m = 4$ genes, dates of birth are known (*via* the generations) as are ploidies and observed genotypes. The goal is to apply our model on this population and to put the results into perspective, compared with the true genealogy $\mathcal{T}^0$ which is represented on the left of Figure 5.

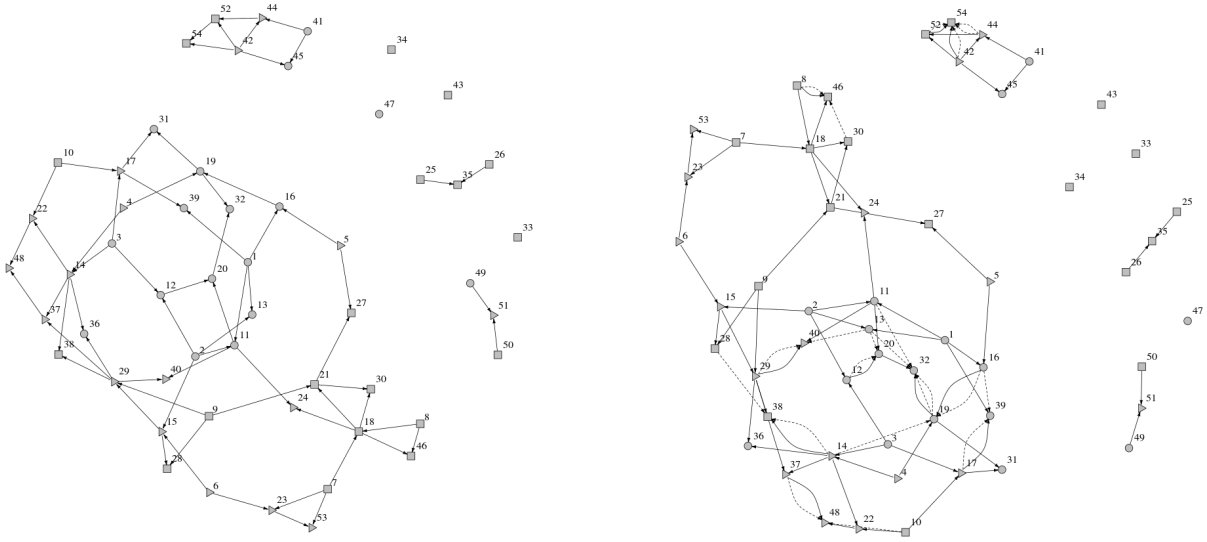


FIGURE 5. True genealogy $\mathcal{T}^0$ of the simulated population, on the left. Superposition of all genealogies of the simulated population found by the model, on the right.

4.1.1. *Family trees and most likely genealogy.* All genealogies found by the model have been superposed on the right of Figure 5, that is, the full content of $G(e)$ given in (2.10) for each $e \in \mathcal{P}$. Similarly, we have also added in Figure 6 the genealogical graph of the population as it is defined in (2.18), highlighting the pairwise potential relationships. We can first verify that the ancestors (individuals from 1 to 10) are only parents. On the one hand, we observe that the true genealogy is included in the graph, illustrating thereby the effectiveness of the exploratory algorithm. One can also notice, on the other hand, that some wrong links have been detected. We should however indicate that a wrong link is not an *impossible* link, for the reader can check that dotted arcs correspond to compatible crosses. Consider as an example the link $\{(14, 28) \mapsto 38\}$ appearing in Figure 5 but absent from the true genealogy. We have $x(14) = 3$, so $\widehat{g}_2(14) = \{160, 170, 180\} = g_2(14)$. Similarly, with $x(28) = 4$ and $x(38) = 4$, $\widehat{g}_2(28) = \{210, 290\}$ can correspond to $g_2(28) = \{210, 290, 290, 290\}$ and $\widehat{g}_2(38) = \{160, 170, 290\}$ to $g_2(38) = \{160, 170, 290, 290\}$. Through pattern ($P_5$), a genealogical link is possible on the signal 2 and we easily check that the same conclusion holds on each signal. This is an illustration of the fact that, from a practical point of view – namely, with an *unknown* true genealogy – it is preferable to produce a set of possible genealogies instead of a single one. Afterwards, the accumulation of genes enables pruning of the trees, step by step, to reinforce the remaining branches. To support this argument, Figure 7 shows on its left the family tree $\mathcal{T}^*$ maximizing the log-likelihood (2.11) in which

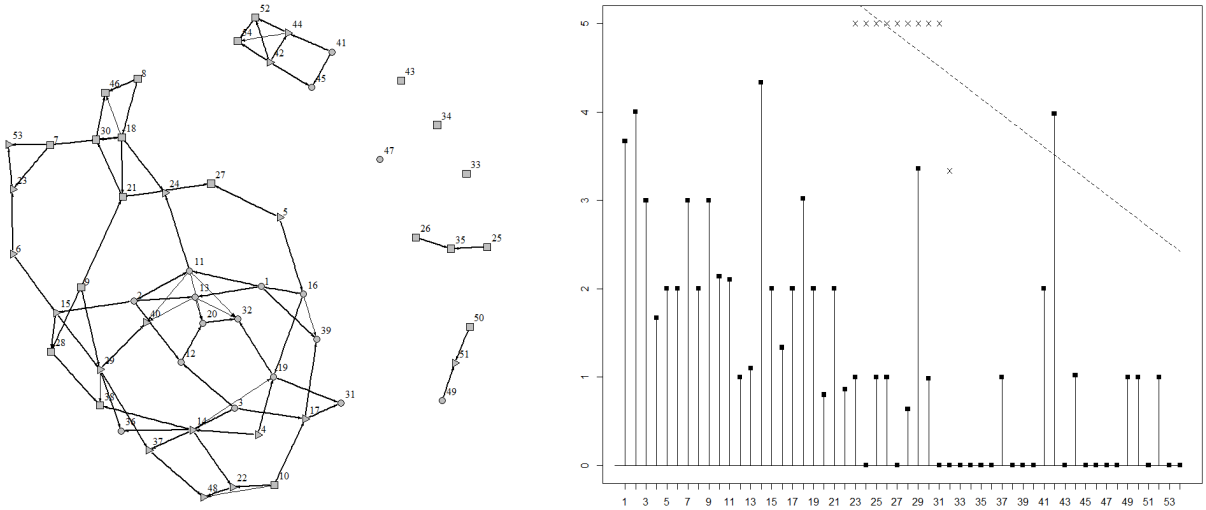


FIGURE 6. Genealogical graph of the simulated population, on the left. The thickness of the links is proportional to their weights in the model. Mean number of offsprings for each individual, on the right. The abscissa displays the individuals $i \in \mathcal{P}$ in chronological order and the ordinate represents the estimated expectation of $N(i)$. The dotted line is the outlier threshold extrapolated from the crosses (the moving window goes through 22 observations). There is 1 probably favored individual.

we observe that the true genealogy was *not* the most likely one, retrospectively. Let us have a look at the differences. The first one is the selection of $\{(14, 28) \mapsto 38\}$ instead of $\{(14, 29) \mapsto 38\}$. Knowing that 28 and 29 both have parents $(9, 15)$, we easily understand their genetic likeness. The second one is interpreted in the same way since $\{(8, 30) \mapsto 46\}$ stands in for $\{(8, 18) \mapsto 46\}$, and since 18 is a parent of 30. For the last two ones, 13 takes the place of 11 in the true connections $\{(11, 12) \mapsto 20\}$ and $\{(11, 29) \mapsto 40\}$, 11 and 13 having the same parents. To be precise, in the latter example each link leads to the same probability and the maximum of likelihood is not unique (in which case the algorithm chooses one solution at random). On this dataset, we get

$$\ell(\mathcal{T}^*) \approx -3.052 \ > \ -7.616 \approx \ell(\mathcal{T}^0).$$

Even so, wrong links maximizing the log-likelihood are usually relevant. In this example, the wrong parents detected are in fact close relatives of true parents. To sum up the results of this simulation, amongst the 45 potential triplets that form the full genealogies, 34 are true and 11 are wrong, but all true links are correctly retrieved. In the maximum likelihood genealogy, one can find from 30 to 32 true links and from 2 to 4 wrong links. The two links that can either be true or wrong have equal probabilities, as it has just been detailed. Even if it is of lesser interest on a simulation, Figure 6 also contains the estimated expectations of the number of offsprings in the population, on the basis of all genealogies with no threshold ($\pi_{\min} = 0$). The individual 42 appears as favored and, indeed, one can check that it has 4 offsprings in the true genealogy whereas it belongs to generation 5. In terms of mean error between the estimated number of offsprings $\widehat{\mathbb{E}}[N(i)]$ and the number of offsprings $n^*(i)$ in

the maximum likelihood genealogy,

$$\frac{1}{n}\sum_{i\in\mathcal{P}}\left|\widehat{\mathbb{E}}[N(i)]-n^*(i)\right|\approx 8.52\times 10^{-2}\quad\text{and}\quad\frac{1}{n}\sum_{i\in\mathcal{P}}\left(\widehat{\mathbb{E}}[N(i)]-n^*(i)\right)^2\approx 5.19\times 10^{-2}.$$

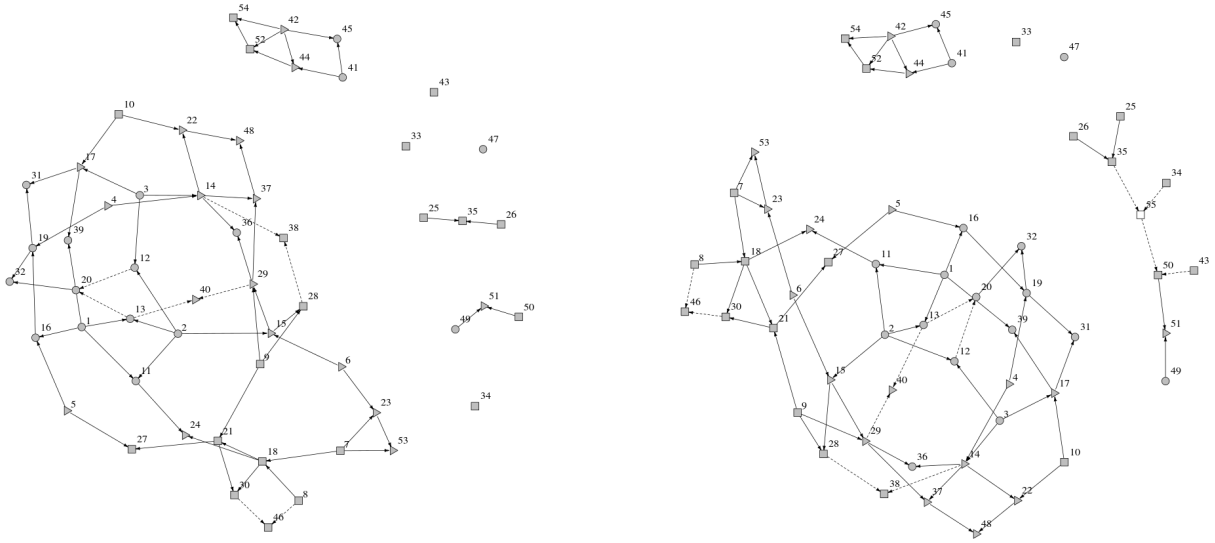


FIGURE 7. Genealogy $\mathcal{T}^*$ maximizing the log-likelihood of the simulated population found by the model, on the left. There are 8 connected components. Genealogy $\mathcal{T}_1$ maximizing the log-likelihood of the simulated population enhanced with one individual (55) found by the model, on the right. There are 5 connected components.

4.1.2. *Missing links.* We now look for missing links, following the algorithm described at the end of Section 3 with $n_v = 3$. Compared with the most likely tree $\mathcal{T}^*$ on the population $\mathcal{P}$, the largest increase of our penalized criterion ℓ^* given by (2.19) is reached by adding the tetraploid $g_1(55) = \{200, 200, 200, 200\}$, $g_2(55) = \{270, 270, 270, 270\}$, $g_3(55) = \{370, 370, 370, 370\}$ and $g_4(55) = \{410, 410, 520, 520\}$, respectively for the 4 genes, as a member of generation 5. We obtain the genealogy on the right of Figure 7. From 8 connected components in $\mathcal{T}^*$, only 5 remain in the maximum likelihood tree $\mathcal{T}_1$ on the population enhanced with the individual 55 having this precise genotype. Thus its role as a missing link is clearly highlighted and that explains the reason why it has been privileged, even if $\ell(\mathcal{T}_1) \approx -3.106$ has decreased compared to $\ell(\mathcal{T}^*) \approx -3.052$. A second loop of the algorithm generates the tetraploid having $g_1(56) = \{10, 10, 200, 200\}$, $g_2(56) = \{130, 130, 380, 380\}$, $g_3(56) = \{210, 210, 370, 370\}$ and $g_4(56) = \{430, 520, 520, 520\}$ on its 4 genes, in generation 3. Only 4 connected components remain, but the log-likelihood is now $\ell(\mathcal{T}_2) \approx -3.482$. The last loop of the algorithm gives a diploid $g_1(57) = \{90, 90\}$, $g_2(57) = \{220, 220\}$, $g_3(57) = \{310, 310\}$ and $g_4(57) = \{510, 510\}$ in generation 5. Only 3 connected components remain while, for this last addition, the log-likelihood is unchanged. Figure 8 depicts $\mathcal{T}_2$ and $\mathcal{T}_3$, respectively on the left and on the right. This simulated example seems to clearly illustrate the operation of the exploratory algorithm, focusing on connected components to build missing links, retrospectively. To support the remarks of Section 3 about the algorithm,

suppose now that the diploid 49 is removed from the dataset. Then, amongst all virtual candidates, a new diploid – say 49* – with genotype $\{240, 240\}$, $\{320, 320\}$, $\{410, 410\}$ and $\{410, 410\}$ appears in generation 6. One can check that this does not correspond to the real 49, but this new genotype allows the cross $\{(49^*, 50) \mapsto 51\}$ with a bigger probability than what actually occurred (precisely, $\frac{1}{2} \times 1 \times \frac{1}{6} \times \frac{1}{12} < \frac{1}{2} \times 1 \times \frac{1}{6} \times \frac{1}{6}$). From this point of view, the algorithm is consistent since there is no way we can retrieve the true allele 510 instead, not spread elsewhere. However, if the diploid 1 is removed from the dataset, then, because it is involved in numerous relationships and because it is heterozygous in most cases, a unique individual playing the same roles is not recovered. For example, on signal 1 and 4, alleles 20 and 310 are needed for $\{(1, 2) \mapsto 13\}$ whereas 10 and 320 are needed for $\{(1, 2) \mapsto 11\}$. The algorithm suggests an individual $\{20, 20\}$ and $\{310, 310\}$ and another one $\{10, 10\}$ and $\{320, 320\}$ on these signals, because they maximize the likelihood of the crossbreedings with 2 to produce 11 and 13. In the end, all genetic information is retrieved but, to be improved, the process should also mix the candidates beforehand, considering $\{10, 20\}$ and $\{310, 320\}$ in this case, as we have mentioned it in the enhancements.

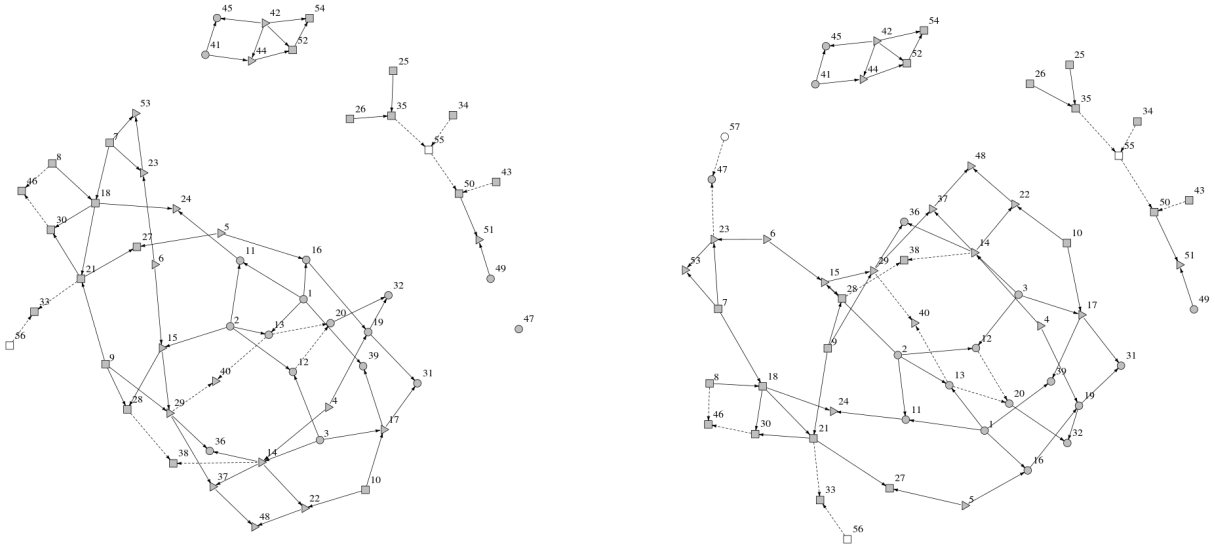


FIGURE 8. Genealogy $\mathcal{T}_2$ maximizing the log-likelihood of the simulated population enhanced with two individuals (55 and 56) found by the model, on the left. There are 4 connected components. Genealogy $\mathcal{T}_3$ maximizing the log-likelihood of the simulated population enhanced with three individuals (55, 56 and 57) found by the model, on the right. There are 3 connected components.

4.2. **On a rose bushes population.** To conclude the study, we are now going to launch our model on a subpopulation of rose bushes collected on the basis of $m = 4$ genes. We start by giving some explanations about the experimental gathering of the data. Among molecular markers, microsatellite markers are still a reference for pedigree reconstruction because they are highly multiallelic codominant markers [9]. After Polymerase Chain Reaction (PCR), amplified fragments are generally separated by capillary electrophoresis. According to their size, amplified fragments are detected at a given time of the electrophoresis and are depicted

as a peak in the electrophoregram, whose area varies according to the intensity of the signal. Thus, a statistical treatment of the four signals of the individual i gives the observed genotypes $\widehat{g}(i)$. To deal with allelic multiplicity, theoretical ratios between peak intensities could be used to determine the relative number of copies of each allele in polyploids [6]. Unfortunately this strategy is very difficult to apply, especially because signal intensity is also dependent on amplification competition between alleles during PCR. Therefore, in most cases electrophoregrams are generally interpreted as presence or absence of alleles [5]. This is also our approach in this article but considering all possibilities of multiplicity, for which we have seen in the previous sections how our model enables building and probabilizing of $g(i)$ from $\widehat{g}(i)$. An example of signal is shown in Figure 9. In addition we must not forget that a calibration of the equipment is needed, for practical purposes. In concrete terms, the abscissa of the signals is made of decimal values, which is clearly incompatible with what it is supposed to highlight, namely some *base pairs*. Hence we take rounded values, and an offset of ± 1 for each allele has to be considered. This is the reason why we decided to switch to criterion (2.5) in the real data analysis.

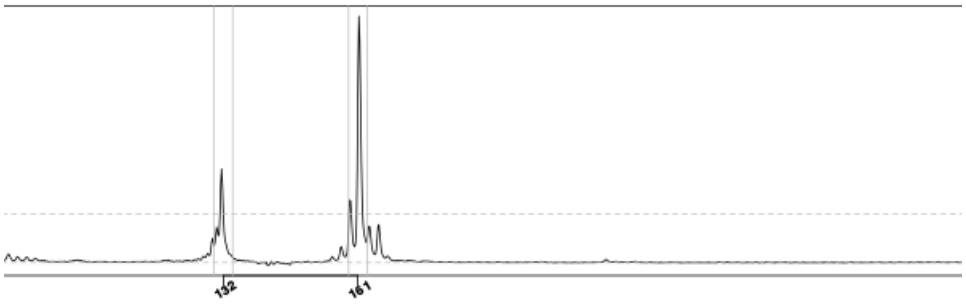


FIGURE 9. Example of signal for a particular microsatellite marker. The individual i is tetraploid and two peaks have been detected. Here $\widehat{g}(i)$ is $\{132, 161\}$ and $g(i)$ is $\{132, 132, 132, 161\}$ with probability π_{31}, $\{132, 132, 161, 161\}$ with probability π_{22} and $\{132, 161, 161, 161\}$ with probability π_{13}. To simplify, scales are deliberately removed.

4.2.1. *Family trees and most likely genealogy.* Now we put aside $n = 116$ rose bushes, selected for the knowledge of their ploidy and for the clarity of their signals, and we look for potential genealogical links among them using the same allelic probabilities as in the simulation study. All genealogies are superposed on Figure 10 together with the genealogical graph on Figure 11 for the threshold probability $\pi_{\min} = 0.2$, a choice that will be justified in the sequel. Even if the graphical representation seems unexploitable, it illustrates the fact that many solutions are conceivable. More than one genealogy maximizes the likelihood, for some links have the same probability. An example of most likely genealogy is given on the left of Figure 12, it contains 35 connected components. Within the largest one, a chain of 5 generations is obtained ($9 \to 56 \to 67 \to 59 \to 47$).

4.2.2. *Missing links.* On the right of Figure 12, one of the most likely genealogies is represented when $n_v = 3$ new individuals suggested by the algorithm of Section 3 are added (117, 118 and 119). Again, their role as missing links and their usefulness to connect separated

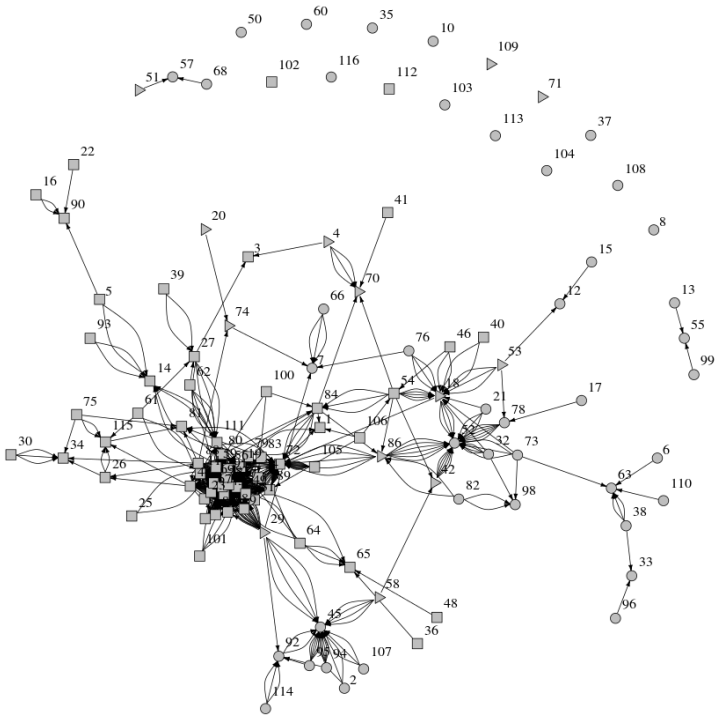


FIGURE 10. Superposition of all genealogies of the rose bushes subpopulation found by the model.

branches of the genealogy are clearly brought to light. Only 32 of them remain, due to the fact that each missing link connects two components. In particular, we can notice the important intercession of 118, plugging the two largest ones.

4.2.3. *Selected individuals.* To look for selected individuals, the estimated probabilities (2.15) and expectations (2.17) are computed for all $i \in \mathcal{P}$ on the basis of a subset of genealogies made of links whose likelihood is greater than $\pi_{\min} = 0.2$. Indeed, since $\mathrm{Card}(\mathbb{G}(\mathcal{P})) > 10^{28}$ the computation with no threshold is infeasible. It appears that with this choice of threshold, $\mathrm{Card}(\mathbb{G}(\mathcal{P}))$ is in the range of 10^6 which is small enough to proceed to computations and large enough to trust the statistical estimations. Figure 13 contains the empirical expectations of all individuals together with an outlier threshold, evaluated as it is explained in the beginning of this section. Each individual having a higher mean number of offsprings is considered as a potential target for the retrospective selection by breeders, there are 6 in this subpopulation. Amongst all individuals, $i = 88$ has, on average, the largest number of offsprings in the population. Figure 14 shows the empirical distribution of $N(88)$. Concretely,

$$\widehat{\mathbb{P}}(N(88) = 5) \approx 0.770, \quad \widehat{\mathbb{P}}(N(88) = 6) \approx 0.230 \quad \text{and} \quad \widehat{\mathbb{E}}[N(88)] \approx 5.230.$$

The last empirical distribution represented is the one of $N(73)$, chosen to illustrate the fact that an individual may have offspring in some genealogies and no offspring in the others. Numerically,

$$\widehat{\mathbb{P}}(N(73) = 0) \approx 0.222, \quad \widehat{\mathbb{P}}(N(73) = 1) \approx 0.444, \quad \widehat{\mathbb{P}}(N(73) = 2) \approx 0.278,$$

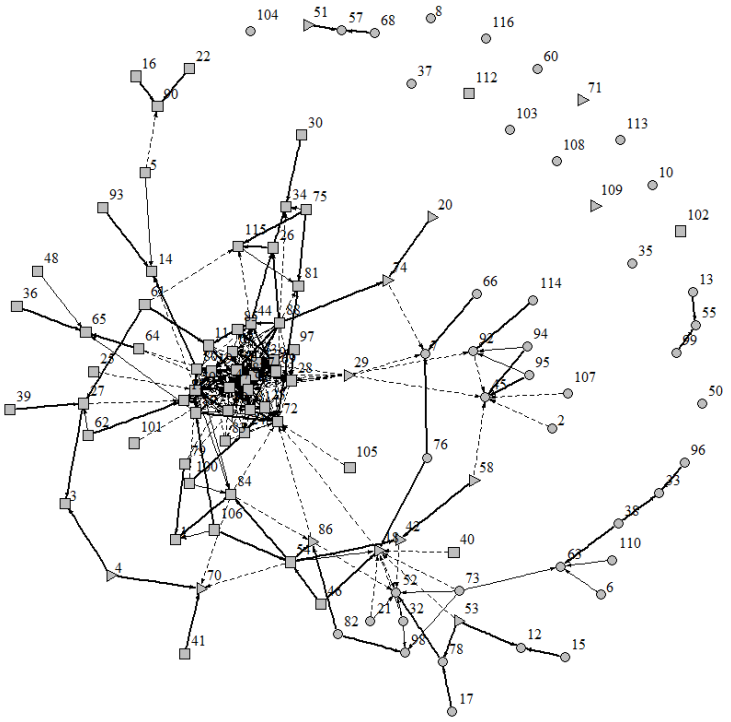


FIGURE 11. Genealogical graph of the rose bushes subpopulation. The thickness of the links is proportional to their weights in the model. The dotted lines correspond to potential links set to zero by the threshold probability.

$$\widehat{\mathbb{P}}(N(73)=3) \approx 0.056 \quad \text{and} \quad \widehat{\mathbb{E}}[N(73)] \approx 1.167.$$

In terms of mean error between the estimated number of offsprings $\widehat{\mathbb{E}}[N(i)]$ and the number of offsprings $n^*(i)$ in the maximum likelihood genealogy,

$$\frac{1}{n}\sum_{i\in\mathcal{P}} \left|\widehat{\mathbb{E}}[N(i)] - n^*(i)\right| \approx 1.21\times 10^{-1} \quad \text{and} \quad \frac{1}{n}\sum_{i\in\mathcal{P}} \left(\widehat{\mathbb{E}}[N(i)] - n^*(i)\right)^2 \approx 7.30\times 10^{-2}.$$

5. Conclusion

To conclude, we would like to draw the attention of the reader to some weaknesses of the model, essentially relying on the allelic multiplicity. Indeed, our choice of considering each potential multiplicity weighted by a probability, instead of selecting a particular one, may lead to contradictions in the genealogy. Suppose for simplification that the most likely genealogy contains the links $\{(p_1,p_2)\mapsto q_1\}$ and $\{(q_1,q_2)\mapsto e\}$ where p_1 is a tetraploid such that $g(p_1)=\{a,a,a,a\}$, and p_2 is a diploid such that $g(p_2)=\{b,b\}$. Both of them are homozygous, so there is no allelic uncertainty derived from their observed genotypes, but $\widehat{g}(q_1)=\{a,b\}$ for the triploid q_1 can only match with $\{(p_1,p_2)\mapsto q_1\}$ in case of $g(q_1)=\{a,a,b\}$. Suppose now that q_2 and e are tetraploids, having $g(q_2)=\{c,c,c,c\}$ and $\widehat{g}(e)=\{b,c\}$, respectively. Then, the link $\{(q_1,q_2)\mapsto e\}$ has a nonzero probability only for $g(q_1)=\{a,b,b\}$. In other words, the most likely genealogy treats q_1 as a link between (p_1,p_2) and e, but at the cost of incompatible allelic combinations. This is a trail for future

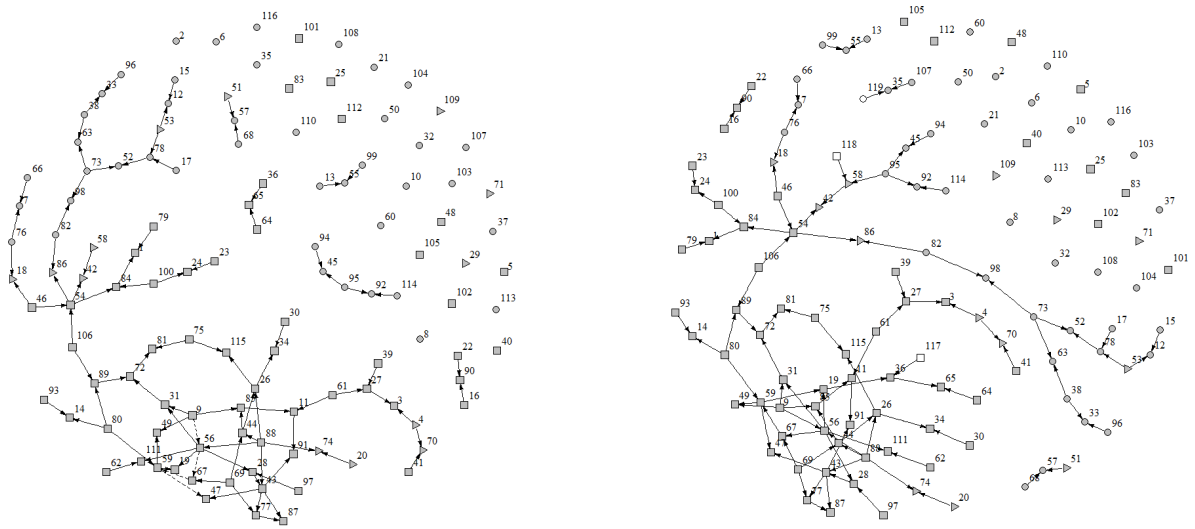


FIGURE 12. Genealogy $\mathcal{T}^*$ maximizing the log-likelihood of the rose bushes subpopulation found by the model (the dotted line highlights a chain of 5 generations), on the left. There are 35 connected components. Genealogy $\mathcal{T}_3$ maximizing the log-likelihood of the rose bushes subpopulation enhanced with three individuals (117, 118 and 119) found by the model, on the right. There are 32 connected components.

improvements of our model, in particular it seems worth considering an algorithm to detect contradictions and to eliminate such trees from the set of genealogies. Another weakness is the estimation of $\pi_{21}, \pi_{12}, \pi_{31}, \ldots$, namely the probabilities of allelic multiplicity. As we have seen in Section 4.2, we lack information to properly evaluate them. An ambitious track could be the generalization of [4], in which the authors establish the well-known *Hardy-Weinberg equilibrium* to deal with heterozygoty in a diploid population. A challenging study will be to characterize this equilibrium in our polyploid population – if it exists – and to determine its degrees of freedom. This additional information will enable us to refine the probabilities of multiplicity, considering that the population has reached its equilibrium. The crossbreeding patterns also have to be enhanced with double reductions and preferential matches, both of them easily treated on a theoretical point of view (dealing with double reductions as rare events of probability $0 < \epsilon \ll 1$ and preferential matches as a lack of uniformity in the gamete production, when computing the probability of the crossbreeding), but difficult to estimate. We have widely discussed the algorithm for missing links and its status of working base which calls for numerous enhancements. Finally, it is important to insist upon the fact that this work is mainly theoretical and that the application of our model on a real population of rose bushes is only relevant in order to show that coherent and interpretable results are obtained. Nevertheless, we cannot draw any conclusion from an empirical study relying on $m = 4$ genes. In-depth experiments will be conducted on more genes, and the comparison of any interesting result with available historical sources will constitute strong arguments to understand the breeders strategies over the past centuries, and also to try to complete the datasets with some lost or missing information.

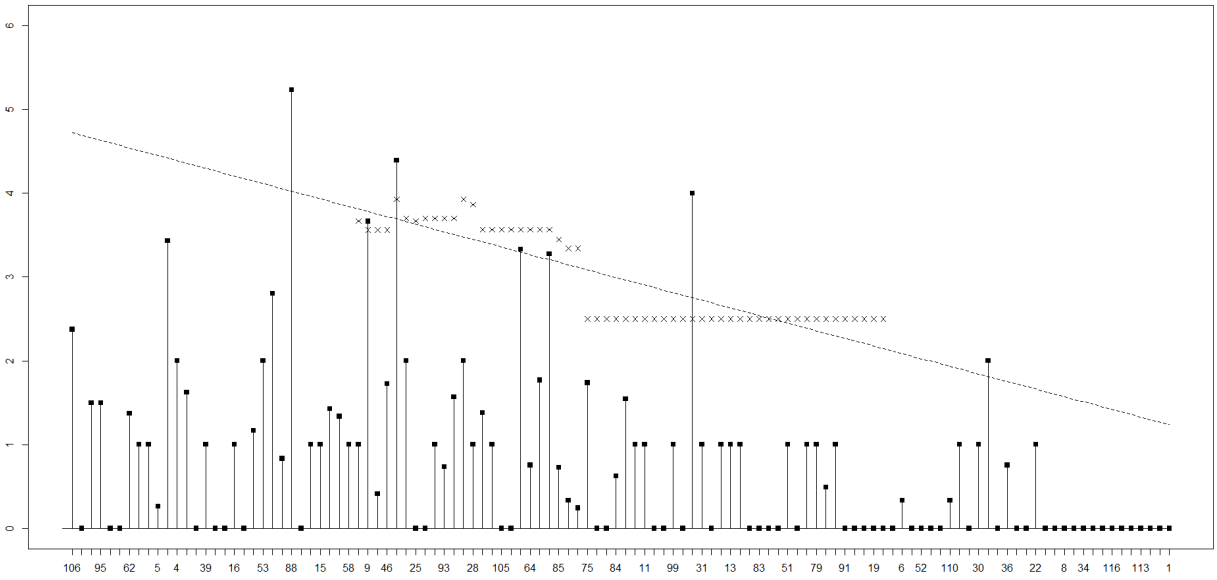


FIGURE 13. Mean number of offsprings for each individual. The abscissa displays the individuals $i \in \mathcal{P}$ in chronological order and the ordinate represents the estimated expectation of $N(i)$. The dotted line is the outlier threshold extrapolated from the crosses (the moving window goes through 30 observations). For readability reasons, the abscissa is not completely filled. There are 6 probably favored individuals.

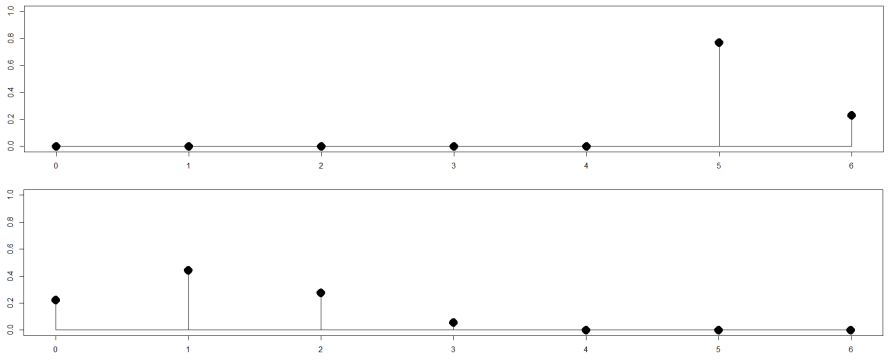


FIGURE 14. Empirical distribution of the random variable $N(88)$, at the top. The abscissa represents the number k of offsprings, the ordinate is the estimated probability associated with the event $\{N(88) = k\}$. At the bottom, empirical distribution of the random variable $N(73)$.

Acknowledgements. This research was conducted in the framework of the regional programme "Objectif Végétal, Research, Education and Innovation in Pays de la Loire", supported by the French Région Pays de la Loire, Angers Loire Métropole and the European Regional Development Fund, in the framework of the PedRo project. Empirical data were obtained thanks to the support of the Région Pays de la Loire in the framework of the FLORHIGE project, by the National Institute of Agricultural Research (INRA) and the

National Natural Science Foundation of China in the framework of the SIFLOR project, and by the French Ministry of Higher Education and Research. The authors also thank the University of Bretagne Loire, Angers Loire Métropole and the University of Angers for their financial support, and the two anonymous reviewers for the numerous constructive comments and suggestions that helped to improve substantially the paper.

Appendix

This Appendix is devoted to the precise description of the simulated population appearing in Section 4.1. All useful information are given in Tables 1, 2 and 3, displaying the composition of the successive generations. For each individual, the columns indicate an identifier i, the ploidy $x(i)$, the observed genotypes $\widehat{g}(i)$ on the four signals, the couple of parents and the reproduction pattern.

Generation 1							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
1	2	10–20	110	210–310	310–320	∅	–
2	2	30–40	130–140	220–230	330	∅	–
3	2	50	150–160	240–250	340	∅	–
4	3	60	170–180–190	260–270	350–360–370	∅	–
5	3	70–80	200	280	380–390–400	∅	–
6	3	90–100–110	210–220	290–300–310	410	∅	–
7	4	120–130–140	230–240–250–260	320–330	420–430–440	∅	–
8	4	150–160–170–180	270–280	340	450	∅	–
9	4	190–200	290–300	350–360–370	460–470–480–490	∅	–
10	4	210–220	310–320	380–390–400	500–510–520	∅	–

Generation 2							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
11	2	10–40	110–130	210–220	320–330	$(1, 2)$	(P_1)
12	2	40–50	140–150	220–250	330–340	$(2, 3)$	(P_1)
13	2	20–40	110–130	210–220	310–330	$(1, 2)$	(P_1)
14	3	50–60	160–170–180	250–270	340–350–370	$(3, 4)$	(P_2)
15	3	40–100–110	140–210	220–290–310	330–410	$(2, 6)$	(P_2)
16	2	20–80	110–200	210–280	320–400	$(1, 5)$	(P_2)
17	3	50–210–220	150–320	240–380–400	340–520	$(3, 10)$	(P_3)
18	4	130–160–180	240–250–270	320–330–340	430–450	$(7, 8)$	(P_6)

Generation 3							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
19	2	20–60	110–180	270–280	350–400	$(4, 16)$	(P_2)
20	2	40	110–150	220	330	$(11, 12)$	(P_1)
21	4	130–180–200	270–290–300	340–350–370	450–480–490	$(9, 18)$	(P_6)
22	3	60–210	180–320	250–390–400	370–520	$(10, 14)$	(P_5)
23	3	90–130–140	220–230–240	300–320	410–420–440	$(6, 7)$	(P_5)
24	3	10–130–160	130–270	210–330–340	330–430–450	$(11, 18)$	(P_3)
25	4	190–200	290–300	350–360–370	410–520	∅	–
26	4	130–160–180	240–250–270	320–330–340	410	∅	–

Table 1. Full description of generations 1, 2 and 3 in the simulated population.

Generation 4							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
27	4	80–200	200–270–300	280–340	380–400–480–490	$(5, 21)$	(P_5)
28	4	40–100–200	210–290	220–310–350–360	330–410–470–490	$(9, 15)$	(P_5)
29	3	100–200	140–290	310–350–370	410–460–490	$(9, 15)$	(P_5)
30	4	130–200	270	330–340–350	450	$(18, 21)$	(P_6)
31	2	20–210	180–320	280–380	400–520	$(17, 19)$	(P_2)
32	2	40–60	110	220–270	330–350	$(19, 20)$	(P_1)
33	4	10–180–200	130–270–380	210–340–370	430–450–520	$\varnothing$	–
34	4	20–90–200	160–270–330	370	520–530–550	$\varnothing$	–
35	4	130–180–200	270–290–300	340–350–370	410	$(25, 26)$	(P_6)

Generation 5							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
36	2	60–100	180–290	270–370	340–490	$(14, 29)$	(P_4)
37	3	50–200	140–160–180	250–270–370	350–370–410	$(14, 29)$	(P_4)
38	4	50–60–100	160–170–290	270–310–350	340–370–410–490	$(14, 29)$	(P_4)
39	2	20–210	110–150	210–380	320–340	$(1, 17)$	(P_2)
40	3	40–200	130–290	210–310–370	330–410–460	$(11, 29)$	(P_2)
41	2	20–110	150–320	260	410–520	$\varnothing$	–
42	3	230	170–390–420	240–340	380–390	$\varnothing$	–
43	4	70–90–100	210–220–270	310–330–340–400	490	$\varnothing$	–

TABLE 2. Full description of generations 4 and 5 in the simulated population.

Generation 6							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
44	3	110–230	170–320–390	240–260–340	390–520	$(41, 42)$	(P_2)
45	2	110–230	320–420	260–340	390–520	$(41, 42)$	(P_2)
46	4	130–150–160	270	330–340	450	$(8, 18)$	(P_6)
47	2	90	220	310–320	410–510	$\varnothing$	–
48	3	50–210	180–320	250–400	410–520	$(22, 37)$	(P_4)
49	2	240	320	410	510–520	$\varnothing$	–
50	4	100–200	270	310–330–370	410–490–520	$\varnothing$	–

Generation 7							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
51	3	200–240	270–320	310–370–410	410–490–520	$(49, 50)$	(P_3)
52	4	230	170–390–420	240–340	390	$(42, 44)$	(P_4)
53	3	130	230–240	320	410–420–440	$(7, 23)$	(P_5)

Generation 8							
i	$x(i)$	$\widehat{g}_1(i)$	$\widehat{g}_2(i)$	$\widehat{g}_3(i)$	$\widehat{g}_4(i)$	Par.	Pat.
54	4	230	170–390–420	240–340	390	$(42, 52)$	(P_5)

TABLE 3. Full description of generations 6, 7 and 8 in the simulated population.

REFERENCES

[1] ACKERMAN, M. W., HAND, B. K., WAPLES, R. K., LUIKART, G., WAPLES, R. S., STEELE, C. A., GARNER, B. A., MCCANE, J., AND CAMPBELL, M. R. Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary applications 10*, 2 (2017), 146–160.

[2] BARKER, M. S., ARRIGO, N., BANIAGA, A. E., LI, Z., AND LEVIN, D. A. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist 210*, 2 (2016), 391–398. 2015-19414.

[3] BOURKE, P. M., ARENS, P., VOORRIPS, R. E., ESSELINK, G. D., KONING-BOUCOIRAN, C. F. S., VAN'T WESTENDE, W. P. C., SANTOS LEONARDO, T., WISSINK, P., ZHENG, C., VAN GEEST, G.,

Visser, R. G. F., Krens, F. A., Smulders, M. J. M., and Maliepaard, C. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *The Plant Journal 90*, 2 (2017), 330–343.

[4] Chaumont, L., Malécot, V., Pymar, R., and Sbai, C. Reconstructing pedigrees using probabilistic analysis of ISSR amplification. *Journal of theoretical biology 412* (2017), 8–16.

[5] Dufresne, F., Stift, M., Vergilino, R., and Mable, B. K. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology 23*, 1 (2014), 40–69.

[6] Esselink, G., Nybom, H., and Vosman, B. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting–peak ratios) method. *Theoretical and Applied Genetics 109*, 2 (2004), 402–408.

[7] Gudin, S., et al. Rose: genetics and breeding. *Plant breeding reviews 17* (2000), 159–190.

[8] Jian, H., Zhang, H., Tang, K., Li, S., Wang, Q., Zhang, T., Qiu, X., and Yan, H. Decaploidy in *Rosa praelucens* byhouwer (Rosaceae) endemic to Zhongdian plateau, Yunnan, China. *Caryologia 63*, 2 (2010), 162–167.

[9] Jones, A. G., and Ardren, W. R. Methods of parentage analysis in natural populations. *Molecular Ecology 12*, 10 (2003), 2511–2523.

[10] Kong, N., Li, Q., Yu, H., and Kong, L.-F. Heritability estimates for growth-related traits in the pacific oyster (*Crassostrea gigas*) using a molecular pedigree. *Aquaculture Research 46*, 2 (2015), 499–508.

[11] Koning-Boucoiran, C. F. S., Gitonga, V. W., Yan, Z., Dolstra, O., van der Linden, C. G., van der Schoot, J., Uenk, G. E., Verlinden, K., Smulders, M. J. M., Krens, F. A., and Maliepaard, C. The mode of inheritance in tetraploid cut roses. *Theoretical and Applied Genetics 125*, 3 (Aug 2012), 591–607.

[12] Lacombe, T., Boursiquot, J.-M., Laucou, V., Di Vecchi-Staraz, M., Péros, J.-P., and This, P. Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* l.). *Theoretical and Applied Genetics 126*, 2 (2013), 401–414.

[13] Liorzou, M., Pernet, A., Li, S., Chastellier, A., Thouroude, T., Michel, G., Malécot, V., Gaillard, S., Brie, C., Foucher, F., Oghina-Pavie, C., Clotault, J., and Grapin, A. Nineteenth century french rose (*Rosa* sp.) germplasm shows a shift over time from a european to an asian genetic background. *Journal of Experimental Botany 67*, 15 (2016), 4711–4725.

[14] Lucena-Perez, M., Soriano, L., López-Bao, J. V., Marmesat, E., Fernández, L., Palomares, F., and Godoy, J. A. Reproductive biology and genealogy in the endangered iberian lynx: Implications for conservation. *Mammalian Biology 89* (2018), 7–13.

[15] Mable, B., Alexandrou, M., and Taylor, M. Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology 284*, 3 (2011), 151–182.

[16] Maia, N., and Venard, P. Cytotaxonomie du genre *Rosa* et origine des rosiers cultivés. 7-20. p. *Travaux sur rosiers de serre Antibes: FNPHP. Cit.: GUDIN, S.(2000): Rose: Genetics and Breeding. Plant Breeding Reviews 17*, 1 (1976), 159–189.

[17] Oghina-Pavie, C. Rose and pear breeding in nineteenth-century france: the practice and science of diversity. In *New Perspectives on the History of Life Sciences and Agriculture*. Springer, 2015, pp. 53–72.

[18] Otto, S. P., and Whitton, J. Polyploid incidence and evolution. *Annual review of genetics 34*, 1 (2000), 401–437.

[19] Raju, D., Namita, K. P. S., Prasad, K., and Janakiram, T. Self-and cross-incompatibility relationship in rose (*Rosa hybrida*) varieties. *Current Horticulture 1*, 2 (2013), 7–9.

[20] Van Huylenbroeck, J., Leus, L., and Van Bockstaele, E. Interploidy crosses in roses: use of triploids. In *1st International Rose Hip Conference* (2005), vol. 690, International Society for Horticultural Science (ISHS), pp. 109–112.

[21] Wang, J., and Scribner, K. T. Parentage and sibship inference from markers in polyploids. *Molecular ecology resources 14*, 3 (2014), 541–553.

Laboratoire angevin de recherche en mathématiques, LAREMA, UMR 6093, CNRS, UNIV Angers, SFR MathSTIC, 2 Bd Lavoisier, 49045 Angers Cedex 01, France.
E-mail address: frederic.proia@univ-angers.fr
E-mail address: fabien.panloup@univ-angers.fr
E-mail address: chiraz.trabelsi@univ-angers.fr

IRHS, Agrocampus-Ouest, INRA, Université d'Angers, SFR 4207 QuaSaV, 49071, Beaucouzé, France.
E-mail address: jeremy.clotault@univ-angers.fr

Bibliographie

Arnold, B., Bomblies, K. et Wakeley, J. (2012). Extending coalescent theory to autotetraploids. *Genetics.*, 192:195–204.

Chaumont, L., Malécot, V., Pymar, R. et Sbai, C. (2017). Reconstructing pedigrees using probabilistic analysis of ISSR amplification. *J. Theor. Biol.*, 412:8–16.

Genolini, C. et Falissard, B. (2010). Kml : k-means for longitudinal data. *Comput. Stat.*, 25:317–328.

Manichaikul, A., Mychaleckyj, J., Rich, S., Daly, K., Sale, M. et Wei-Min, C. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics.*, 26:2867–2873.

Proïa, F., Panloup, F., Trabelsi, C. et Clotault, J. (2019). Probabilistic reconstruction of genealogies for polyploid plant species. *J. Theor. Biol.*, 462:537–551.

Proïa, F., Pernet, A., Thouroude, T., Michel, G. et Clotault, J. (2016). On the characterization of flowering curves using Gaussian mixture models. *J. Theor. Biol.*, 402:75–88.

Wiper, M., Rios Insua, D. et Ruggeri, F. (2001). Mixtures of Gamma distributions with applications. *J. Comput. Graph. Stat.*, 10(3):440–454.