

Optimisation Riemannienne pour l'apprentissage de graphes structurés

Alexandre HIPPERT FERRER¹ Florent BOUCHARD² Ammar MIAN³ Titouan VAYER⁴ Arnaud BRELOY⁵

¹Univ. Gustave Eiffel, LASTIG, ² CNRS, L2S, CentraleSupélec, ³ INRIA, LIP, ENS-Lyon,

⁴Univ. Savoie Mont-blanc, LISTIC, ⁵ Univ. Paris Nanterre, LEME

Résumé – Cet article propose un cadre algorithmique pour l'apprentissage de graphes au travers de l'estimation de matrices de précision parcimonieuses sous des contraintes de structure rang faible sur la matrice de covariance. Le problème est exprimé comme une estimation pénalisée du maximum de vraisemblance d'une distribution elliptique (une généralisation des modèles graphiques gaussiens aux distributions à queue lourde), où la matrice de covariance est optionnellement contrainte à avoir une structure rang faible plus diagonale. La résolution de cette classe de problèmes est ensuite abordée via une méthode d'optimisation riemannienne, où nous tirons parti de géométries des matrices définies positives et des matrices semi-définies positives de rang fixe qui sont bien adaptées aux modèles elliptiques.

Abstract – This paper proposes an algorithmic framework for graph learning through sparse precision matrix estimation under low-rank structural constraints on the covariance matrix. The problem is expressed as penalized maximum likelihood estimation of an elliptical distribution (a generalization of Gaussian graphical models to possibly heavy-tailed distributions), where the covariance matrix is optionally constrained to be structured as low-rank plus diagonal (low-rank factor model). The resolution of this class of problems is then tackled with Riemannian optimization, where we leverage geometries of positive definite matrices and positive semi-definite matrices of fixed rank that are well suited to elliptical models.

1 Introduction

Les modèles graphiques permettent de représenter des structures de dépendances entre des variables (entrées) d'observation multivariées $\mathbf{x} = [x_1, \dots, x_p]^\top \in \mathbb{R}^p$. Concrètement, chaque variable x_q , $q \in \llbracket 1, p \rrbracket$ est représentée par un sommet du graphe, et une arête (q, l) est présente dans le graphe entre deux sommets x_q et x_l si “un lien” existe entre ces variables existe. L'apprentissage de graphe fait alors référence au problème d'inférer la topologie (inconnue) d'un graphe sous-jacent à une série d'observations $\{\mathbf{x}_i\}_{i=1}^n \in (\mathbb{R}^p)^n$.

D'un point de vue statistique le lien en question peut être interprété comme l'existence d'une corrélation conditionnelle : pour les modèles graphiques gaussiens (GGM), on suppose que chaque observation est une réalisation d'un vecteur aléatoire gaussien multivarié centré $\mathbf{x} = [x_1, \dots, x_p]^\top$ avec une matrice de covariance $\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \Sigma$, notée $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ [10]. Le support de la matrice de précision $\Theta = \Sigma^{-1}$ rend alors compte de la structure de dépendance conditionnelle entre les variables, car

$$\text{corr}[x_q x_\ell | \mathbf{x}_{\llbracket 1, p \rrbracket \setminus \{q, \ell\}}] = -\Theta_{q\ell} / \sqrt{\Theta_{qq} \Theta_{\ell\ell}}. \quad (1)$$

Par conséquent, une entrée non nulle $\Theta_{q\ell}$ implique une dépendance conditionnelle entre les variables x_q et x_ℓ , et donc une arête entre les sommets q et ℓ du graphe. Notons parallèlement que cette propriété se généralise aussi à d'autres modèles multivariés [11]. Dans les GGM, l'apprentissage de graphe est donc lié au problème de l'estimation de la matrice de précision Θ à partir d'un ensemble d'observations $\{\mathbf{x}_i\}_{i=1}^n \in (\mathbb{R}^p)^n$. Afin de mettre en évidence une telle structure de corrélation, une approche standard consiste à recourir à l'estimation régularisée du maximum de vraisemblance, *i.e.*, en résolvant :

$$\underset{\Theta \in \mathcal{S}_p^{++}}{\text{maximize}} \quad \log \det(\Theta) - \text{Tr}\{\mathbf{S}\Theta\} - \lambda h(\Theta), \quad (2)$$

où $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$ est la matrice de covariance empirique de l'échantillon, h est une pénalité de régularisation et $\lambda \in \mathbb{R}^+$ est un paramètre de régularisation. La norme ℓ_1 est souvent utilisée comme pénalité afin de promouvoir une structure parcimonieuse dans Θ . Cette formulation est à la base de l'algorithme GLasso [5, 13, 4], qui a donné lieu à de nombreuses généralisations : structure de parcimonie particulière, généralisations à des modèles non-Gaussiens [20], a priori de structures additionnels sur la matrice de covariances [3, 9], etc.

Dans cette continuité, le travail présenté dans cette communication¹ vise à étendre la formulation dans deux directions :

- Nous considérons la généralisation de la fonction de vraisemblance à des modèles elliptiques. Cette famille de distributions inclut en effet le modèle gaussien, mais aussi, de nombreuses autres distributions multivariées à queues lourdes. La dérivation d'algorithmes dans ce cadre général permet alors d'assurer une certaine robustesse à des données aberrantes.
- Nous considérons un modèle à facteur rang faible, *i.e.*, où la matrice de covariance est structurée comme rang faible plus diagonale. Cette structure réduit la dimension du problème d'estimation, ce qui permet *i)* de pouvoir traiter des jeux de données avec un nombre d'échantillons réduit ($n \ll p$); *ii)* de pouvoir développer des algorithmes dont la complexité supporte un passage à l'échelle en grande dimension p .

Ces extensions amènent à formalisation unifiée de problèmes d'optimisation non-triviaux (contraintes complexes sur la matrice de covariance, et pénalités s'appliquant sur son inverse). Nous proposons alors un cadre algorithmique pour résoudre cette famille de problème en tirant parti de l'optimisation Riemannienne [1].

¹Cet article synthétise les résultats présentés dans [6]. Il est supporté par l'ANR MASSILIA (ANR-21-CE23-0038-01).

2 Motivations des généralisations

2.1 Distributions elliptiques

Une première limitation des algorithmes tels que formulés dans (1) est leur manque de robustesse face aux valeurs aberrantes, ou aux échantillons suivant une distribution à queue lourde. Ceci est une conséquence de l'hypothèse gaussienne, qui n'est pas adaptée aux distributions empiriques de telles données. Une solution possible est de considérer une famille plus large de distributions : dans ce contexte, les distributions elliptiques [7] offrent une alternative connue pour fournir des estimateurs robustes de la matrice de covariance [12, 19], qui à été utilisée avec succès pour étendre les modèles graphiques [20, 2]. Un vecteur suit une loi elliptique centrée de matrice de dispersion Σ et de générateur de densité g , noté $\mathbf{x} \sim \mathcal{ES}(\mathbf{0}, \Sigma, g)$, si sa densité de probabilité est de la forme

$$f(\mathbf{x}) \propto \det(\Sigma)^{-1/2} g(\mathbf{x}_i^\top \Sigma^{-1} \mathbf{x}_i), \quad (3)$$

qui donne la fonction de log-vraisemblance

$$\mathcal{L}(\Sigma) \propto \frac{1}{n} \sum_{i=1}^n \rho(\mathbf{x}_i^\top \Sigma^{-1} \mathbf{x}_i) + \frac{1}{2} \log |\Sigma| + \text{const.} \quad (4)$$

pour l'échantillon $\{\mathbf{x}_i\}_{i=1}^n$, où $\rho(t) = -\log g(t)$. On remarque que $g(t) = \exp(-t/2)$ correspond au cas gaussien. Cependant, le générateur de densité g permet une plus grande flexibilité, et notamment d'englober de nombreuses distributions multivariées à queue lourde. Parmi les choix usuels, les distributions elliptiques comprennent la loi de Student (t -distribution) de degré de libertés $\nu > 2$, obtenue avec $g(t) = (1 + t/\nu)^{-\frac{\nu+p}{2}}$. Pour cette distribution, le paramètre ν pilote la lourdeur de la queue de distribution. Enfin, il est particulièrement intéressant de noter que dès que l'on suppose une distribution à queue lourde, il n'est pas nécessaire de connaître précisément g pour assurer de bonnes performances sur l'ensemble de la famille des lois elliptiques [14].

2.2 Modèles de covariance structurés

Une deuxième limitation de (1) est que les structures potentielles de la matrice de covariance ne sont pas prises en compte. Ceci peut notamment poser problème quand le support de l'échantillon est faible ($n \simeq p$, ou $n < p$). Dans ce contexte, une approche consiste à imposer une structure rang faible. Cette structure vient de l'hypothèse sous-jacente d'un modèle linéaire de faible dimension, i.e., les données sont supposées suivre $\mathbf{x} = \mathbf{W}\mathbf{s} + \epsilon$, où $\mathbf{W} \in \mathbb{R}^{p \times k}$ est une matrice k facteurs, et $\mathbf{s} \in \mathbb{R}^k$ sont $\epsilon \in \mathbb{R}^p$ sont des variables aléatoires indépendantes. La matrice de covariance résultante est de la forme $\Sigma \triangleq \mathbf{H} + \Psi$, où $\mathbf{H} = \mathbf{W}\mathbb{E}[\mathbf{s}\mathbf{s}^\top]\mathbf{W}^\top$ appartient à l'ensemble des matrices semi-définies positives de rang k , noté $\mathcal{S}_{p,k}^+ = \{\Sigma \in \mathcal{S}_p, \Sigma \succcurlyeq \mathbf{0}, \text{rank}(\Sigma) = k\}$. Ce modèle est, par exemple, celui de l'analyse probabiliste en composantes principales, ou l'on suppose $\Psi \propto \mathbf{I}_p$ [18]. Dans cet article, nous nous concentrerons le cas général [16, 8], où $\Psi \in \mathcal{D}_p^{++}$, où $\mathcal{D}_p^{++} = \{\Sigma = \text{diag}(\mathbf{d}), \mathbf{d} \in \mathbb{R}_{*}^p\}$ représente l'espace des matrices diagonales définies positives. Ainsi, la matrice de covariance appartient à l'espace des matrices de rang- k plus diagonales, noté par

$$\mathcal{M}_{p,k} = \left\{ \Sigma = \mathbf{H} + \Psi, \mathbf{H} \in \mathcal{S}_{p,k}^+, \Psi \in \mathcal{D}_p^{++} \right\}. \quad (5)$$

Par ailleurs, cette paramétrisation réduit la dimension du problème d'estimation de $p(p+1)/2$ à $p(k+1) - k(k-3)/2$, ce qui explique qu'il soit souvent utilisé dans des contextes de grande dimension p .

3 Formalisation du problème

Nous formalisons le problème d'optimisation suivant :

$$\begin{aligned} & \underset{\Sigma \in \mathcal{S}_p^{++}}{\text{minimize}} && \mathcal{L}(\Sigma) + \lambda h(\Sigma) \\ & \text{subject to} && \Sigma \in \mathcal{M}_{p,k} \end{aligned} \quad (6)$$

où \mathcal{L} est la fonction de log-vraisemblance (4), $\mathcal{M}_{p,k}$ est l'espace des matrices défini en (5). La fonction h est une pénalité qui promeut une structure parcimonieuse du graphe, et $\lambda \in \mathbb{R}^+$ est un paramètre de régularisation. Le choix de la pénalité dépend du type de structures recherchées et de l'application : dans ce travail, nous nous concentrons sur une pénalité élément-par-élément appliquée sur $\Theta = \Sigma^{-1}$, définie comme :

$$h(\Sigma) = \sum_{q \neq \ell} \phi([\Sigma^{-1}]_{q\ell}), \quad (7)$$

qui promeut une parcimonie sans motif particulier. Il est important de préciser que le cadre algorithmique de résolution considéré requiert que h soit régulière (à minima continue et différentiable en tout point). Nous considérons ici une approximation régulière de la norme ℓ_1 , définie par

$$\phi(t) = \varepsilon \log(\cosh(t/\varepsilon)), \quad (8)$$

avec $\varepsilon > 0$. La norme ℓ_1 correspond au cas limite $\lim_{\varepsilon \rightarrow 0} \phi(t)$.

En résumé, le problème (6) rend compte d'une structure de rang faible des données dans un problème d'apprentissage robuste de graphe. On peut distinguer quatre cas principaux, et les acronymes correspondants :

- GGM/GGFM : le modèle gaussien graphique de rang faible (GGFM) est obtenu avec une log-vraisemblance gaussienne, i.e., avec $g(t) = \exp(-t/2)$ en (4). Le modèle gaussien graphique classique (GGM) de (1) correspond au sans la contrainte $\Sigma \in \mathcal{M}_{p,k}$.
- EGM/EGFM : le modèle elliptique graphique de rang faible (EGFM) est obtenu dans le cas général où \mathcal{L} est une fonction de log-vraisemblance d'une loi elliptique. De même, relaxer la contrainte $\Sigma \in \mathcal{M}_{p,k}$ donne un modèle elliptique graphique non structuré (EGM).

4 Optimisation riemannienne

4.1 Cadre théorique

Le problème général (6) s'avère non-trivial car il implique des contraintes complexes sur la matrice de covariance, ainsi qu'une pénalité s'appliquant sur son inverse. Nous proposons une approche de résolution mettant à profit le formalisme de l'optimisation riemannienne. Tout d'abord, remarquons que la variable appartient à une variété, i.e., $\Sigma \in \mathcal{M}$ (ici \mathcal{M} représente $\mathcal{M}_{p,k}$ en (5), ou \mathcal{S}_p^{++} pour le cas non contraint), qui possède un espace tangent $T_{\Sigma}\mathcal{M}$ en chaque point Σ . Définir une métrique, i.e., un produit scalaire $\langle \cdot, \cdot \rangle_{\Sigma}$ (régulier en Σ)

en chaque espace M , donne une variété riemannienne. Ce point de vue permet d'utiliser de nombreux outils d'optimisation [1]. Par exemple, la méthode de la descente de gradient riemannienne s'exprime

$$t_{i+1} = R^M_i(t_i(r^M f(t_i))) \quad (9)$$

ou deux outils sont conditionnés par la métrique choisie (et la géométrie induite sur la variété): Le gradient riemannien r^M , qui est l'unique vecteur de M satisfaisant

$$\langle r^M f(t); i^M = Df(t)[\cdot] \rangle; \quad (10)$$

où $Df(t)[\cdot]$ est la dérivée directionnelle de f dans la direction \cdot . ii) Une rétraction R^M , qui est un mapping local de T^M vers M . L'ensemble des outils en question pour les deux variétés $M_{p;k}$ et S_p^{++} , munies d'un choix de métrique approprié aux distributions elliptiques, sont détaillées dans [6]. Les sections suivantes donnent les formules nécessaires à l'implémentation de la descente de gradient riemannienne pour ces deux cas. Le pas de gradient est sélectionné par backtracking et les algorithmes sont initialisés avec la matrice de covariance empirique (son EVD pour la partie rang faible).

4.2 Outils pour S_p^{++}

La variété S_p^{++} est un ouvert de S_p , l'espace tangent à S_p^{++} peut donc être identifié à S_p pour tout $t \in S_p^{++}$. La métrique choisie est

$$h; i^{S_p^{++}} = \text{tr}(\cdot^T \cdot); \quad (11)$$

Ce qui induit le gradient riemannien

$$r^{S_p^{++}} f(t) = \text{sym}(r^{E_f}(t)); \quad (12)$$

et la rétraction du second ordre

$$R^{S_p^{++}}(t) = t + \frac{1}{2} r^{S_p^{++}}(t)^2; \quad (13)$$

4.3 Outils pour $M_{p;k}$

Une matrice $2 \times M_{p;k}$ peut s'exprimer $V V^T + S_k D_p$, avec $V \in \mathbb{R}^{2 \times p}$, $S_k \in \mathbb{R}^{2 \times 2}$ et $D_p \in \mathbb{R}^{2 \times 2}$. Définissons $N_{p;k} = S_k^{++} D_p^{++}$ et

$$i^{N_{p;k}} : (V; S_k; D_p) \mapsto \text{tr}(V V^T + S_k D_p); \quad (14)$$

Ainsi, $M_{p;k} = i^{N_{p;k}}$, ce qui sera utilisé comme représentation pour résoudre (6). L'espace tangent à $M_{p;k}$ est obtenu par agrégation des espaces $T_{p;k}$, $T_{S_k^{++}}$ et $T_{D_p^{++}}$, i.e.,

$$T_{N_{p;k}} = f(V; S_k; D_p) \in \mathbb{R}^{2 \times p} \times \mathbb{R}^{2 \times 2} \times \mathbb{R}^{2 \times 2}; \quad (15)$$

La métrique choisie est

$$h; i^{N_{p;k}} = \text{tr}(\frac{1}{2} V V^T + S_k D_p) + \text{tr}(S_k) + \text{tr}(D_p); \quad (16)$$

²Notons aussi que cette paramétrisation contient une classe d'équivalence, et que nous considérons le quotient $N_{p;k} = O_k$ dans les détails de [6]

Ce qui induit une formule de gradient riemannien $r^{N_{p;k}} \bar{f}(t)$ sur $N_{p;k}$ comme fonction du gradient euclidien $r^{E_f}(t)$.

$$r^{N_{p;k}} \bar{f}(t) = (G_V - V G_V^T V; G; \text{ddiag}(G)); \quad (17)$$

Notons aussi que $r^{E_f}(t)$ peut s'obtenir comme fonction du gradient euclidien $r^{E_f}(t)$ via

$$r^{E_f}(t) = (2 r^{E_f}(t) V; V^T r^{E_f}(t) V; \text{ddiag}(r^{E_f}(t))) ; \quad (18)$$

La rétraction

$$R^{N_{p;k}}(t) = (u f(V + v); \frac{1}{2} \text{tr}(V V^T) + \frac{1}{2} \text{tr}(S_k) + \frac{1}{2} \text{tr}(D_p)); \quad (19)$$

où $u f(\cdot)$ renvoie le facteur orthogonal de la décomposition polaire.

5 Applications sur données GNSS

Nous présentons une application des méthodes proposées à des données de déplacement de la surface terrestre collectées par un réseau de récepteurs GNSS de l'observatoire volcanologique du Piton de la Fournaise (OVVP-IPGP), situé à La Réunion. Le réseau présenté est composé de 22 récepteurs enregistrant les déplacements verticaux quotidiens de janvier 2014 à mars 2017 [7], avec un total de $n = 1106$ observations. Durant cette période, les déplacements verticaux induits par les éruptions volcaniques ont été enregistrés par le réseau de capteurs, incluant parfois des mouvements rapides. En fonction de leur position spatiale, certains récepteurs peuvent se déplacer dans une direction particulière (vers le haut ou vers le bas), laissant possible l'extraction d'une structure de corrélations spatiales entre récepteurs.

Les résultats de l'apprentissage des graphes par les différents algorithmes sont présentés en Figure 1, où chaque récepteur est un nœud du graphe. Les graphes obtenus révèlent trois composantes de deux situés à l'ouest (récepteurs 'GITG', 'FREG', etc.) et à l'est ('BORG', 'FJAG', etc.) du cratère sommital, et un s'étendant des altitudes inférieures jusqu'au bord de mer ('CASG', 'TRCG', etc.). Nous appelons ces composantes respectivement west et east low. Comme décrit dans [15], les quatre éruptions de 2015 (février, mai, juillet et août) sont caractérisées par des motifs de déplacement asymétriques par rapport aux fissures éruptives nord-sud qui s'étendent légèrement vers l'ouest du cratère sommital. Il est intéressant de noter que cela correspond à la séparation entre les composantes du graphes west et east qui est mieux mise en évidence par les algorithmes basés sur les modèles par facteur, en particulier EGFM. Cela illustre également l'intérêt de la prise en compte d'une distribution non Gaussienne pour les données considérées. La composante low correspond aux récepteurs dont le déplacement est faible ou nul. Notons que le récepteur 'PRAG' est également inclus dans ce groupe, probablement parce qu'il n'a pas enregistré de mouvement significatif au cours de cette période. En EGGM, EGM et StGL conduisent à des résultats similaires, mais avec des connexions erronées entre le côté ouest du cratère et les récepteurs du bord de mer (e.g. de 'BORG' à 'CASG' pour StGL, et 'PRAG' connecté à certains récepteurs des composantes west/ouest).

(a) StGL ($m = 0.66$)

(b) GGM ($m = 0.51$)

(c) EGM ($m = 0.59$)

(d) GGFM ($m = 0.61$)

(e) EGFM ($m = 0.6$)

FIGURE 1 : Graphes appris sur les données GNSS par StGL et les méthodes proposées (m est la modularité). StGL requiert un nombre de clusters dans le graphe (ici fixé à 3). Pour GGFM and EGFM, le rang est $k = 4$. Le nombre de degrés de liberté est fixé à $\nu = 5$. L'étoile rouge représente le sommet du cratère ©WEBBOBS/IPGP.

Références

- [1] P.-A. ABSIL, R. MAHONY et R. SEPULCHRE : *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, USA, 2008.
- [2] Jose Vinicius de MIRANDA CARDOSO, Jiayi YING et Daniel PALOMAR : Graphical models in heavy-tailed markets. *Advances in Neural Information Processing Systems*, 34:19989–20001, 2021.
- [3] Hilmi E EGILMEZ, Eduardo PAVEZ et Antonio ORTEGA : Graph learning from data under laplacian and structural constraints. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):825–841, 2017.
- [4] Salar FATTAHI et Somayeh SOJUDI : Graphical lasso and thresholding : Equivalence and closed-form solutions. *Journal of machine learning research*, 2019.
- [5] Jerome FRIEDMAN, Trevor HASTIE et Robert TIBSHIRANI : Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- [6] Alexandre HIPPERT-FERRER, Florent BOUCHARD, Ammar MIAN, Titouan VAYER et Arnaud BRELOY : Learning graphical factor models with riemannian optimization. *arXiv preprint arXiv :2210.11950*, 2022.
- [7] Fang KAI-TAI et Zhang YAO-TING : *Generalized multivariate analysis*, volume 19. Science Press Beijing and Springer-Verlag, Berlin, 1990.
- [8] Koulik KHAMARU et Rahul MAZUMDER : Computation of the maximum likelihood estimator in low-rank factor analysis. *Mathematical Programming*, 176(1):279–310, 2019.
- [9] Sandeep KUMAR, Jiayi YING, José Vinicius de MIRANDA CARDOSO et Daniel P PALOMAR : A unified framework for structured graph learning via spectral constraints. *J. Mach. Learn. Res.*, 21(22):1–60, 2020.
- [10] Steffen L LAURITZEN : *Graphical models*, volume 17. Clarendon Press, 1996.
- [11] Han LIU, John LAFFERTY et Larry WASSERMAN : The nonparanormal : Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10(10), 2009.
- [12] Ricardo Antonio MARONNA : Robust m-estimators of multivariate location and scatter. *The annals of statistics*, pages 51–67, 1976.
- [13] Rahul MAZUMDER et Trevor HASTIE : The graphical lasso : New insights and alternatives. *Electronic journal of statistics*, 6:2125, 2012.
- [14] Esa OLLILA, David E TYLER, Visa KOIVUNEN et H Vincent POOR : Complex elliptically symmetric distributions : Survey, new results and applications. *IEEE Transactions on signal processing*, 60(11):5597–5625, 2012.
- [15] Aline PELTIER, Jean-Luc FROGER, Nicolas VILLENEUVE et Thibault CATRY : Assessing the reliability and consistency of InSAR and GNSS data for retrieving 3D-displacement rapid changes, the example of the 2015 Piton de la Fournaise eruptions. *Journal of Volcanology and Geothermal Research*, 344:106–120, 2017.
- [16] Donald ROBERTSON et James SYMONS : Maximum likelihood factor analysis with rank-deficient sample covariance matrices. *Journal of Multivariate Analysis*, 98(4):813–828, 2007.
- [17] Delphine SMITTARELLO, Valérie CAYOL, Virginie PINEL, Aline PELTIER, J-L FROGER et Valérie FERRAZZINI : Magma propagation at Piton de la Fournaise from joint inversion of InSAR and GNSS. *Journal of Geophysical Research : Solid Earth*, 124(2):1361–1387, 2019.
- [18] Michael E TIPPING et Christopher M BISHOP : Probabilistic principal component analysis. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 61(3):611–622, 1999.
- [19] Yoav WALD, Nofar NOY, Gal ELIDAN et Ami WIESEL : Globally optimal learning for structured elliptical losses. *Advances in Neural Information Processing Systems*, 32, 2019.
- [20] Teng ZHANG, Ami WIESEL et Maria Sabrina GRECO : Multivariate generalized gaussian distribution : Convexity and graphical models. *IEEE Transactions on Signal Processing*, 61(16):4141–4148, 2013.