Journal of Data Science 0 (0), 1–4
July 2022
DOI: 0000
Statistical Data Science

Supplementary Material for "FROSTY: A High-dimensional Scale-free Bayesian Network Learning Method"

Joshua Bang[1,*] and Sang-Yun Oh[2]
[1]Department of Statistics and Applied Probability, University of California, Santa Barbara, Santa Barbara, CA, USA
[2]Lawrence Berkeley National Lab, Berkeley, CA, USA

1 Proofs

1.1 Preliminaries

Recall that our loss function parametrized with $L \in \mathbb{L}_p$ is

$$\ell(S;L) = \text{trace}(SLL^T) - 2\log|L|$$

and its gradient is

$$\frac{\partial}{\partial L}\ell(S;L) = 2SL - 2(L^{-1})^T.$$

Let $\mathcal{P}(\mathbb{S}_p)$ be the set of all probability distributions supported on $\mathbb{S}_p$ and $S_1,\dots,S_n$ be an independent and identically distributed random sample from S coming from a distribution on $\mathcal{P}(\mathbb{S}_p)$. Now, we restate the following two propositions, suitable to our context, that are the key ingredients for proofs of the theorems.

Proposition 1.1 (Proposition C.2 of Cisneros et al. (2020))**.** *For $\gamma \geq 0$ and loss functions $\ell(S';L)$ that are upper semi-continuous in $S' \in \mathbb{S}_p$ for each $L \in \mathbb{L}_p$, let*

$$\phi_\gamma(S_i;L) = \sup_{S\in\mathbb{S}_p}\{\ell(S;L) - \gamma c(S,S_i)\}. \tag{1}$$

Then,

$$\sup_{\mathcal{P}:D_c(\mathcal{P},\mathcal{P}_n)\leq\lambda} E_{\mathcal{P}}[\ell(S;L)] = \min_{\gamma\geq 0}\left\{\gamma\lambda + \frac{1}{n}\sum_{i=1}^n \phi_\gamma(S_i;L)\right\}. \tag{2}$$

Proposition 1.2 (Proposition C.1 of Cisneros et al. (2020))**.** *Consider $L \in \mathbb{L}_p$. Let $h(\cdot,L)$ be Borel measurable. Also, suppose that $\mathbb{0}_{p\times p}$ lies in the interior of the convex hull of $\{h(S',C) : S' \in \mathbb{S}_p\}$. Then,*

$$R_n(L) = \sup_{\Lambda\in\mathbb{S}_p}\left\{-\frac{1}{n}\sum_{i=1}^n \sup_{S\in\mathbb{S}_p}\{\text{trace}(\Lambda^T h(S;L)) - c(S,S_i)\}\right\}. \tag{3}$$

*Corresponding author. Email: joshuaybang@gmail.com.

1.2 Proof of Theorem 3.1

Proof. Using the duality from Proposition 1.1, we can rewrite the left-hand side of the equation (4) from the theorem

$$\begin{aligned}\min_{L\in\mathbb{L}_p}\sup_{\mathcal{P}:D_c(\mathcal{P},\mathcal{P}_n)\le\lambda} E_{\mathcal{P}}[\ell(S;L)] &= \min_{L\in\mathbb{L}_p}\left[\min_{\gamma\ge 0}\left\{\gamma\lambda+\frac{1}{n}\sum_{i=1}^{n}\phi_\gamma(S_i;L)\right\}\right]\\ &= \min_{L\in\mathbb{L}_p}\left[\min_{\gamma\ge 0}\left\{\gamma\lambda+\frac{1}{n}\sum_{i=1}^{n}\left(\sup_{S\in\mathbb{S}_p}\{\ell(S;L)-\gamma c(S,S_i)\}\right)\right\}\right]. \quad (4)\end{aligned}$$

Let $\Delta = S - S_i$. Then,

$$\begin{aligned}\sup_{S\in\mathbb{S}_p}\{\ell(S;L)-\gamma c(S,S_i)\} &= \sup_{S\in\mathbb{S}_p}\{\mathrm{trace}(SLL^T)-2\log|L|-\gamma\|\mathrm{vec}(S)-\mathrm{vec}(S_i)\|_\infty\}\\ &= \sup_{\Delta\in\mathbb{S}_p}\{\mathrm{trace}((\Delta+S_i)LL^T)-\gamma\|\mathrm{vec}(\Delta)\|_\infty\}-2\log|L|\\ &= \sup_{\Delta\in\mathbb{S}_p}\{\mathrm{trace}(\Delta LL^T)-\gamma\|\mathrm{vec}(\Delta)\|_\infty\}+\mathrm{trace}(S_iLL^T)-2\log|L|\\ &= \sup_{\Delta\in\mathcal{M}(L)}\{\|\mathrm{vec}(\Delta)\|_\infty\|\mathrm{vec}(LL^T)\|_1-\gamma\|\mathrm{vec}(\Delta)\|_\infty\}+\mathrm{trace}(S_iLL^T)-2\log|L|\\ &= \sup_{\Delta\in\mathcal{M}(L)}\{\|\mathrm{vec}(\Delta)\|_\infty(\|\mathrm{vec}(LL^T)\|_1-\gamma)\}+\mathrm{trace}(S_iLL^T)-2\log|L|\end{aligned}\quad (5)$$

where $\mathcal{M}(L)=\{\Delta\in\mathbb{S}_p:\mathrm{trace}(\Delta LL^T)>0, |(\Delta)_{ij}|^q=|(LL^T)_{ij}|^p\}$ with $\frac{1}{p}+\frac{1}{q}=1$, which is the if and only if condition for Holder's inequality to hold tightly (Steele, 2004). Since this condition also holds for the limiting case, we choose $p=1$ and $q=\infty$.

Notice that if $\gamma<\|\mathrm{vec}(LL^T)\|_1$, the supremum term in the last line of (5) can be arbitrarily large. Since we have the minimum over $\gamma\ge 0$ in (4), we must take $\gamma\ge\|\mathrm{vec}(LL^T)\|_1$, which leads to the supremum term to be zero. Then, we have

$$\begin{aligned}\min_{L\in\mathbb{L}_p}\sup_{\mathcal{P}:D_c(\mathcal{P},\mathcal{P}_n)\le\lambda} E_{\mathcal{P}}[\ell(S;L)] &= \min_{L\in\mathbb{L}_p}\left[\min_{\gamma\ge\|\mathrm{vec}(LL^T)\|_1}\left\{\gamma\lambda+\frac{1}{n}\sum_{i=1}^{n}(\mathrm{trace}(S_iLL^T)-2\log|L|)\right\}\right]\\ &= \min_{L\in\mathbb{L}_p}\left\{\mathrm{trace}(SLL^T)-2\log|L|+\lambda\|\mathrm{vec}(LL^T)\|_1\right\}.\end{aligned}$$

□

1.3 Proof of Theorem 3.2

Proof. Consider the RWP function with $h(S;L)=2SL-2(L^{-1})^T$. Applying 1.2, we have

$$\begin{aligned}R_n(L) &= \sup_{\Lambda\in\mathbb{S}_p}\left\{-\frac{1}{n}\sum_{i=1}^{n}\sup_{S\in\mathbb{S}_p}\{\mathrm{trace}(\Lambda^T h(S;L))-c(S,S_i)\}\right\}\\ &= \sup_{\Lambda\in\mathbb{S}_p}\left\{-\frac{1}{n}\sum_{i=1}^{n}\sup_{S\in\mathbb{S}_p}\{\mathrm{trace}(2\Lambda^T(SL-(L^{-1})^T))-\|\mathrm{vec}(S)-\mathrm{vec}(S_i)\|_\infty\}\right\}.\end{aligned}$$

Let $\Delta = S - S_i$. Then, following the procedure of (5) in the proof of theorem 3.1, we obtain

$$
\begin{aligned}
&\sup_{S\in\mathbb{S}_p}\{\text{trace}(2\Lambda^T(SL-(L^{-1})^T)) - \|\text{vec}(S)-\text{vec}(S_i)\|_\infty\} \\
&= \sup_{\Delta\in\mathbb{S}_p}\{\text{trace}(2\Lambda^T((\Delta+S_i)L-(L^{-1})^T)) - \|\text{vec}(\Delta)\|_\infty\} \\
&= \sup_{\Delta\in\mathbb{S}_p}\{\text{trace}(2\Lambda^T\Delta L) - \|\text{vec}(\Delta)\|_\infty\} + \text{trace}(2\Lambda^T(S_iL-(L^{-1})^T)) \\
&= \sup_{\Delta\in\mathcal{M}(\Lambda)}\{\|\text{vec}(\Delta)\|_\infty\|\text{vec}(2L\Lambda^T)\|_1 - \|\text{vec}(\Delta)\|_\infty\} + \text{trace}(2\Lambda^T(S_iL-(L^{-1})^T)) \\
&= \sup_{\Delta\in\mathcal{M}(\Lambda)}\{\|\text{vec}(\Delta)\|_\infty(\|\text{vec}(2L\Lambda^T)\|_1 - 1)\} + \text{trace}(2\Lambda^T(S_iL-(L^{-1})^T)). \qquad (6)
\end{aligned}
$$

Similarly, we must consider $\|\text{vec}(2L\Lambda^T)\|_1 \le 1$ to prevent the supremum term in the last line of (6) from being taken arbitrarily large. Then, we have

$$
\begin{aligned}
R_n(L) &= \sup_{\Lambda\in\mathbb{S}_p:\|\text{vec}(2L\Lambda^T)\|_1\le 1}\left\{-\frac{1}{n}\sum_{i=1}^n \text{trace}(2\Lambda^T(S_iL-(L^{-1})^T))\right\} \\
&= \sup_{\Lambda\in\mathbb{S}_p:\|\text{vec}(2L\Lambda^T)\|_1\le 1}\left\{-\text{trace}(2\Lambda^T(SL-(L^{-1})^T))\right\} \\
&= \sup_{\Lambda\in\mathbb{S}_p:\|\text{vec}(2L\Lambda^T)\|_1\le 1}\left\{-\text{trace}((SLL^{-1}-(L^{-1})^TL^{-1})2L\Lambda^T)\right\} \\
&= \sup_{\Lambda\in\mathbb{S}_p:\|\text{vec}(2L\Lambda^T)\|_1\le 1}\left\{-\text{trace}((S-(LL^T)^{-1})2L\Lambda^T)\right\} \\
&= \sup_{\Lambda\in\mathbb{S}_p:\|\text{vec}(2L\Lambda^T)\|_1\le 1}\left\{\text{vec}(2L\Lambda^T)^T\text{vec}(S-(LL^T)^{-1})\right\} \\
&= \|\text{vec}(S-(LL^T)^{-1})\|_\infty.
\end{aligned}
$$

Finally, since

$$
\{L\in\mathcal{C}_n(\lambda)\} = \{\mathcal{O}(L)\cap\{\mathcal{P}: D_c(\mathcal{P},\mathcal{P}_n)\le\lambda\}\ne\emptyset\} = \{R_n(L)\le\lambda\},
$$

one can, given a $(1-\alpha)$-confidence level, optimally choose the ambiguity size λ by

$$
\begin{aligned}
\lambda_\alpha &= \inf\{\lambda>0: \mathbb{P}_0(L\in\mathcal{C}_n(\lambda))\ge 1-\alpha\} \\
&= \inf\{\lambda>0: \mathbb{P}_0(R_n(L)\le\lambda)\ge 1-\alpha\} \\
&= \inf\{\lambda>0: \mathbb{P}_0(\|\text{vec}(S-(LL^T)^{-1})\|_\infty\le\lambda)\ge 1-\alpha\}.
\end{aligned}
$$

□

1.4 Proof of Lemma 3.3

Proof.

$$
\begin{aligned}
R_n(\Theta) &= \|\text{vec}(S-\Theta^{-1})\|_\infty \\
&= \|\text{vec}(S-(P_\pi^TL_\pi L_\pi^TP_\pi)^{-1})\|_\infty \\
&= \|\text{vec}(S-P_\pi^T(L_\pi L_\pi^T)^{-1}P_\pi)\|_\infty \\
&= \|\text{vec}(P_\pi^T[P_\pi SP_\pi^T-(L_\pi L_\pi^T)^{-1}]P_\pi)\|_\infty \\
&= \|\text{vec}(P_\pi^T[S_\pi-(L_\pi L_\pi^T)^{-1}]P_\pi)\|_\infty \\
&= \|\text{vec}(S_\pi-(L_\pi L_\pi^T)^{-1})\|_\infty = R_n(L_\pi).
\end{aligned}
$$

□

References

Cisneros P, Petersen A, Oh SY (2020). Distributionally robust formulation and model selection for the graphical lasso. In: *International Conference on Artificial Intelligence and Statistics*, 756–765.

Steele JM (2004). *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*. Cambridge University Press.