# Learning the piece-wise constant graph structure of a varying Ising model: Appendix

**Batiste Le Bars** [1] **Pierre Humbert** [1] **Argyris Kalogeratos** [1] **Nicolas Vayatis** [1]

## Additional figures and results

### Comparison between $\ell_2$- and $\ell_1$-norms

In Fig. 1, we illustrate the main difference in using an $\ell_2$- or alternatively a $\ell_1$-norm in the fused penalty of our objective function. The figure illustrates well the problem of $\ell_1$-norm: by penalizing each dimension independently, this norm easily leads to parameter vectors that have some non-zero dimensions, making the piece-wise constant assumption more difficult to recover. On the contrary, the $\ell_2$-norm avoids this problem and hence enforces the whole consecutive parameter vectors to be equal.

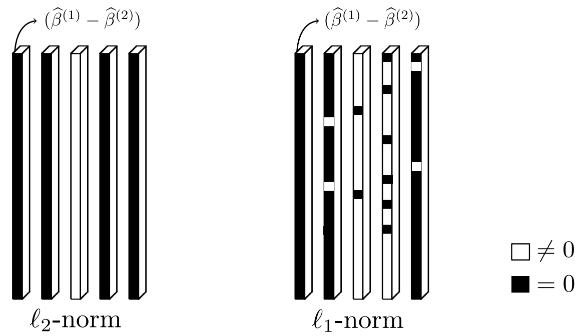


**Figure 1:** Comparison of the learned parameter vectors when using either $\ell_2$- or $\ell_1$-norm in the fused penalty. White squares indicates dimensions at which the two consecutive parameter vectors are different. Black squares where they are equal. The presence of at least one white square indicates a change-point.

### Another real-world experiment

In this section, we evaluate the goodness of graph learning with TVI-FL on the Sigfox IoT dataset (Le Bars and Kalogeratos, 2019) (available at: http://kalogeratos.com/the-sigfox-iot-dataset). The dataset contains activity recorded on a telecommunication network, where each observation corresponds to a message that was locally broadcasted by one device and has been received by a subset of the 34 monitored antennas. Each data vector is binary and indicates which antennas has received the message or not (received $= 1$, not received $= 0$). The dataset contains all the messages received by the antennas, on a daily basis over a period of five months, resulting in $n = 120$ timestamps. According to the authors, one antenna is working poorly after the 30-th timestamp. In the following experiment, we select this antenna along with the 19 geographically closest others, and we select randomly $n_i = 200$ messages at each timestamp. The learned graphs with TVI-FL at timestamps $i = 0$ (before the antenna's malfunction) and $i = 60$ (after the antenna's malfunction) are displayed in Fig. 2, where only the positive edges are drawn.

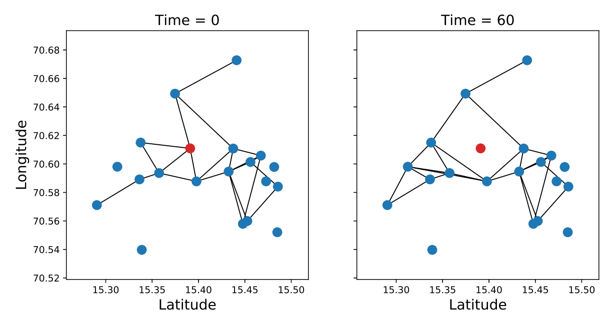


**Figure 2:** Learned graphs for Sigfox dataset before and after the anomaly recorded at the red antenna.

The goodness-of-fit of our method can be corroborated by the observations: 1) The learned graphs are in agreement with the spatial distribution of the antennas: nearby antennas are more likely to be connected as they have high chance to receive the same messages; 2) The problematic antenna lost edges after its malfunction. Again, this is as expected since a poorly working antenna would receive less messages, implying a decreased correlation with its neighbors.

### A more complete table of results

For completeness, in Tab. 1 we complete the table of the main text of the paper with an additional comparative

[1]Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-91190 Gif-sur-Yvette, France. Correspondence to: Batiste Le Bars <batiste.lebars@cmla.ens-cachan.fr>.

*Proceedings of the $37^{th}$ International Conference on Machine Learning*, Online, PMLR 119, 2020. 

method, namely the one that estimates a graph at each time-stamp ($\lambda_1 = 0$).

## Technical proofs

### Main results

In the following, we recall and prove the main results given in the paper. The proofs uses in many situations the different lemmas given next.

**Lemma 1.** *(Optimality Conditions) A matrix $\hat{\beta}$ is optimal for our problem iff there exists a collection of subgradient vectors $\{\hat{z}^{(i)}\}_{i=2}^{n}$ and $\{\hat{y}^{(i)}\}_{i=1}^{n}$, with $\hat{z}^{(i)} \in \partial\|\widehat{\beta}^{(i)} - \widehat{\beta}^{(i-1)}\|_2$ and $\hat{y}^{(i)} \in \partial\|\widehat{\beta}^{(i)}\|_1$, such that $\forall k = 1, ..., n$ we have:*

$$\sum_{i=k}^{n} x_{\backslash a}^{(i)} \left\{ \tanh\Big(\widehat{\beta}^{(i)\intercal} x_{\backslash a}^{(i)}\Big)\Big) - \tanh\Big(\omega_a^{(i)\intercal} x_{\backslash a}^{(i)}\Big)\Big) \right\} - \sum_{i=k}^{n} x_{\backslash a}^{(i)} \left\{ x_a^{(i)} - \mathbb{E}_{\Omega^{(i)}}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right] \right\} + \lambda_1 \hat{z}^{(k)} + \lambda_2 \sum_{i=k}^{n} \hat{y}^{(i)} = \mathbf{0}_{p-1}, \quad (1)$$

*where* $\tanh$ *is the hyperbolic tangent function,* $\mathbf{0}_{p-1}$ *is the zero vector of size* $p-1$, $\hat{z}^{(1)} = \mathbf{0}_{p-1}$, *and*

$$\hat{z}^{(i)} = \begin{cases} \frac{\widehat{\beta}^{(i)} - \widehat{\beta}^{(i-1)}}{\|\widehat{\beta}^{(i)} - \widehat{\beta}^{(i-1)}\|_2} & \textit{if } \widehat{\beta}^{(i)} - \widehat{\beta}^{(i-1)} \neq 0, \\ \in \mathcal{B}_2(0,1) & \textit{otherwise}; \end{cases}$$

$$\hat{y}^{(i)} = \begin{cases} \text{sign}(\widehat{\beta}^{(i)}) & \textit{if } x \neq 0, \\ \in \mathcal{B}_1(0,1) & \textit{otherwise}. \end{cases}$$

*Proof.* Let us first introduce the following change of variables:

$$\gamma^{(i)} = \begin{cases} \beta^{(i)} & \text{if} \quad i = 1, \\ \beta^{(i)} - \beta^{(i-1)} & \text{otherwise}. \end{cases}$$

Thus $\beta^{(i)} = \sum_{l=1}^{i} \gamma^{(l)}$, which leads to a change in the objective function (4) of the main paper:

$$\{\hat{\gamma}^{(i)}\}_{i=1}^{n} = \underset{\gamma \in \mathbb{R}^{p-1 \times n}}{\text{argmin}} \sum_{i=1}^{n} \log\left\{ \exp\left(\sum_{l=1}^{i} \gamma^{(l)\intercal} x_{\backslash a}^{(i)}\right) + \exp\left(-\sum_{l=1}^{i} \gamma^{(l)\intercal} x_{\backslash a}^{(i)}\right) \right\} - \sum_{i=1}^{n} x_a^{(i)} \sum_{l=1}^{i} \gamma^{(l)\intercal} x_{\backslash a}^{(i)} + \lambda_1 \sum_{i=2}^{n} \|\gamma^{(i)}\|_2 + \lambda_2 \sum_{i=1}^{n} \|\sum_{l=1}^{i} \gamma^{(l)}\|_1. \quad (2)$$

This problem is convex, thus a necessary and sufficient condition for $\{\hat{\gamma}^{(i)}\}_{i=1}^{n}$ to be a solution is that for all $k = 1, ..., n$, the $(p-1)$-dimensional zero-vector $\mathbf{0}$, belongs to the subdifferential of (2), taken with respect to $\gamma^{(k)}$:

$$\mathbf{0} \in \sum_{i=k}^{n} x_{\backslash a}^{(i)} \left(\tanh\left(\sum_{l=1}^{i} \hat{\gamma}^{(l)\intercal} x_{\backslash a}^{(i)}\right) - x_a^{(i)}\right) + \lambda_1 \partial\|\hat{\gamma}^{(k)}\|_2 + \lambda_2 \sum_{i=k}^{n} \partial\|\sum_{l=1}^{i} \hat{\gamma}^{(l)}\|_1.$$

Recall that

$$\partial\|x\|_2 = \begin{cases} \left\{\frac{x}{\|x\|_2}\right\} & \text{if} \quad x \neq 0 \\ \mathcal{B}_2(0,1) & \text{otherwise}; \end{cases}$$

$$\partial\|x\|_1 = \begin{cases} \{\text{sign}(x)\} & \text{if} \quad x \neq 0 \\ \mathcal{B}_1(0,1) & \text{otherwise} \end{cases}.$$

Reapplying the change of variable, we obtain:

$$\mathbf{0} = \sum_{i=k}^{n} x_{\backslash a}^{(i)} \left(\tanh\left(\hat{\beta}^{(i)\intercal} x_{\backslash a}^{(i)}\right) - x_a^{(i)}\right) + \lambda_1 \hat{z}^{(k)} + \lambda_2 \sum_{i=k}^{n} \hat{y}^{(i)}.$$

Noting that $\mathbb{E}_{\Omega^{(i)}}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right] = \tanh\Big(\omega_a^{(i)\intercal} x_{\backslash a}^{(i)}\Big)\Big)$, we obtain the final result. ☐

**Theorem 1.** *(Change-point consistency) Let $\{x_i\}_{i=1}^{n}$ be a sequence of observations drawn from the piece-wise constant Ising model presented in Sec. 2. Suppose (A1-A3) hold, and assume that $\lambda_1 \asymp \lambda_2 = \mathcal{O}(\sqrt{\log(n)/n})$. Let $\{\delta_n\}_{n \geq 1}$ be a non-increasing sequence that converges to 0, and such that $\forall n > 0$, $\Delta_{\min} \geq n\delta_n$, with $n\delta_n \to +\infty$. Assume further that $(i)$ $\frac{\lambda_1}{n\delta_n \xi_{\min}} \to 0$, $(ii)$ $\frac{\sqrt{p-1}\lambda_2}{\xi_{\min}} \to 0$, and $(iii)$ $\frac{\sqrt{p\log(n)}}{\xi_{\min}\sqrt{n\delta_n}} \to 0$. Then, if the correct number of change-points are estimated, we have $\widehat{D} = D$ and:*

$$\mathbb{P}(\max_{j=1,...,D} |\hat{T}_j - T_j| \leq n\delta_n) \underset{n \to \infty}{\longrightarrow} 1. \quad (3)$$

*Proof.* The proof follows the steps given in (Harchaoui and Lévy-Leduc, 2010; Kolar and Xing, 2012; Gibberd and Roy, 2017). First of all, thanks to the union bound,

$$\mathbb{P}(\max_{j=1,...,D} |\hat{T}_j - T_j| > n\delta_n) \leq \sum_{j=1}^{D} \mathbb{P}(|\hat{T}_j - T_j| > n\delta_n),$$

thus it suffices to show for each $j = 1, ..., D$, that $\mathbb{P}(|\hat{T}_j - T_j| > n\delta_n) \to 0$. We denote by $A_{n,j}$ the event $\Big\{|\hat{T}_j - T_j| > n\delta_n\Big\}$.

Similarly to (Kolar and Xing, 2012), we first consider the good case where we assume that the event $C_n = \Big\{|\hat{T}_j - T_j| < \frac{\Delta_{\min}}{2}\Big\}$ occurs.

| Degree | Observations per timestamp | Method | AIC $h$-score ↓ | AIC $F_1$-score ↑ | AIC $\widehat{D}$ | AUC $h$-score ↓ | AUC $F_1$-score ↑ | AUC $\widehat{D}$ |
|---|---|---|---|---|---|---|---|---|
| $d=2$ | $n^{(i)}=4$ | TVI-FL | **0.046 ± (0.024)** | **0.694 ± (0.103)** | 7.400 ± (3.137) | 0.221 ± (0.035) | **0.876 ± (0.030)** | 26.100 ± (7.739) |
| | | Tesla | 0.106 ± (0.087) | 0.649 ± (0.190) | 12.700 ± (7.682) | **0.184 ± (0.051)** | 0.841 ± (0.041) | 25.100 ± (4.784) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.342 ± (0.007) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.342 ± (0.000) | 99.000 ± (0.000) |
| | $n^{(i)}=6$ | TVI-FL | **0.129 ± (0.058)** | **0.816 ± (0.073)** | 9.700 ± (2.759) | **0.147 ± (0.071)** | **0.875 ± (0.027)** | 15.300 ± (3.378) |
| | | Tesla | 0.178 ± (0.130) | 0.748 ± (0.167) | 12.900 ± (5.540) | 0.164 ± (0.062) | 0.841 ± (0.048) | 19.000 ± (2.530) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.407 ± (0.010) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.407 ± (0.000) | 99.000 ± (0.000) |
| | $n^{(i)}=8$ | TVI-FL | **0.082 ± (0.081)** | 0.833 ± (0.095) | 7.400 ± (3.040) | **0.099 ± (0.073)** | **0.891 ± (0.024)** | 11.000 ± (3.873) |
| | | Tesla | 0.124 ± (0.071) | **0.846 ± (0.047)** | 13.600 ± (2.010) | 0.178 ± (0.066) | 0.853 ± (0.039) | 14.700 ± (3.348) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.449 ± (0.009) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.449 ± (0.000) | 99.000 ± (0.000) |
| $d=3$ | $n^{(i)}=4$ | TVI-FL | **0.080 ± (0.069)** | **0.563 ± (0.089)** | 7.000 ± (2.683) | **0.204 ± (0.035)** | **0.734 ± (0.024)** | 23.100 ± (6.715) |
| | | Tesla | 0.278 ± (0.319) | 0.353 ± (0.072) | 3.200 ± (2.891) | 0.208 ± (0.029) | 0.611 ± (0.041) | 29.200 ± (3.187) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.366 ± (0.010) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.366 ± (0.000) | 99.000 ± (0.000) |
| | $n^{(i)}=6$ | TVI-FL | **0.055 ± (0.064)** | **0.617 ± (0.161)** | 6.300 ± (3.494) | **0.130 ± (0.051)** | **0.743 ± (0.034)** | 12.800 ± (2.821) |
| | | Tesla | 0.302 ± (0.241) | 0.346 ± (0.060) | 2.000 ± (1.183) | 0.173 ± (0.044) | 0.616 ± (0.041) | 22.600 ± (2.245) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.391 ± (0.014) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.391 ± (0.000) | 99.000 ± (0.000) |
| | $n^{(i)}=8$ | TVI-FL | **0.091 ± (0.073)** | **0.714 ± (0.130)** | 8.000 ± (2.530) | **0.127 ± (0.073)** | **0.764 ± (0.032)** | 10.400 ± (2.154) |
| | | Tesla | 0.311 ± (0.231) | 0.361 ± (0.098) | 2.600 ± (2.615) | 0.162 ± (0.052) | 0.633 ± (0.045) | 18.700 ± (3.716) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.410 ± (0.015) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.410 ± (0.000) | 99.000 ± (0.000) |
| $d=4$ | $n^{(i)}=4$ | TVI-FL | **0.101 ± (0.082)** | **0.453 ± (0.111)** | 6.500 ± (3.324) | **0.232 ± (0.026)** | **0.644 ± (0.041)** | 29.400 ± (4.317) |
| | | Tesla | 0.444 ± (0.273) | 0.347 ± (0.044) | 2.875 ± (1.900) | 0.234 ± (0.017) | 0.518 ± (0.046) | 34.625 ± (1.654) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.388 ± (0.005) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.388 ± (0.000) | 99.000 ± (0.000) |
| | $n^{(i)}=6$ | TVI-FL | **0.099 ± (0.064)** | **0.501 ± (0.130)** | 5.667 ± (2.309) | **0.183 ± (0.044)** | **0.664 ± (0.041)** | 16.778 ± (3.258) |
| | | Tesla | 0.258 ± (0.236) | 0.355 ± (0.035) | 2.500 ± (1.118) | 0.215 ± (0.032) | 0.503 ± (0.040) | 26.000 ± (4.472) |
| | | Static | 0.290 ± (0.000) | 0.390 ± (0.007) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.390 ± (0.000) | 99.000 ± (0.000) |
| | $n^{(i)}=8$ | TVI-FL | **0.077 ± (0.076)** | **0.528 ± (0.158)** | 5.556 ± (3.624) | **0.169 ± (0.064)** | **0.678 ± (0.049)** | 12.444 ± (4.524) |
| | | Tesla | 0.251 ± (0.230) | 0.357 ± (0.044) | 2.625 ± (0.696) | 0.219 ± (0.027) | 0.518 ± (0.054) | 24.000 ± (2.398) |
| | | $\lambda_1=0$ | 0.290 ± (0.000) | 0.385 ± (0.007) | 99.000 ± (0.000) | 0.290 ± (0.000) | 0.385 ± (0.000) | 99.000 ± (0.000) |

**Table 1:** Results for the model with the lowest AIC, and that with the highest AUC. The average ± (std) of the metrics is reported. Compared to the table provided in the main text, here an additional comparative method is mentioned, namely the one that estimates a graph at each timestamp ($\lambda_1 = 0$).

### Bounding the good case

For each $j = 1, ..., D$, we show that $\mathbb{P}(A_{n,j} \cap C_n) \longrightarrow 0$. In particular, we suppose that $\hat{T}_j \leq T_j$ as the proof for $\hat{T}_j \geq T_j$ will be the same by symmetry.
Applying Lemma 1 with $k = \hat{T}_j$ and $k = T_j$, subtracting one with the other and applying the $\ell_2$-norm, we obtain:

$$0 = \Bigg\| \sum_{i=\hat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \left\{\tanh\left(\widehat{\beta}^{(i)\top} x_{\backslash a}^{(i)}\right) - \tanh\left(\omega_a^{(i)\top} x_{\backslash a}^{(i)}\right)\right\} - \sum_{i=\hat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \left\{x_a^{(i)} - \mathbb{E}_{\Omega^{(i)}}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right]\right\} + \lambda_1(\hat{z}^{(\widehat{T}_j)} - \hat{z}^{(T_j)}) + \lambda_2 \sum_{i=\hat{T}_j}^{T_j-1} \hat{y}^{(i)} \Bigg\|_2$$

$$\geq \Bigg\| \sum_{i=\hat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \left\{\tanh\left(\widehat{\beta}^{(i)\top} x_{\backslash a}^{(i)}\right) - \tanh\left(\omega_a^{(i)\top} x_{\backslash a}^{(i)}\right)\right\} - \sum_{i=\hat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \left\{x_a^{(i)} - \mathbb{E}_{\Omega^{(i)}}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right]\right\} \Bigg\|_2 - \Bigg\| \lambda_2 \sum_{i=\hat{T}_j}^{T_j-1} \hat{y}^{(i)} \Bigg\|_2 - \left\| \lambda_1(\hat{z}^{(\widehat{T}_j)} - \hat{z}^{(T_j)}) \right\|_2.$$

We have:
$\left\| \lambda_1(\hat{z}^{(\widehat{T}_j)} - \hat{z}^{(T_j)}) \right\|_2 \leq 2\lambda_1$ and
$\left\| \lambda_2 \sum_{i=\hat{T}_j}^{T_j-1} \hat{y}^{(i)} \right\|_2 \leq (T_j - \hat{T}_j)\sqrt{p-1}\lambda_2$.
Furthermore, one may notice that for all $i \in \left\{\hat{T}_j, ..., T_j - 1\right\}$, $\widehat{\beta}^{(i)} = \hat{\theta}_a^{j+1}$ and $\omega_a^{(i)} = \theta_a^j$. Adding and subtracting $\tanh\left((\theta_a^{j+1})^\top x_{\backslash a}^{(i)}\right)$, then applying again the triangle inequality, leads to the following result:

$$2\lambda_1 + (T_j - \hat{T}_j)\sqrt{p-1}\lambda_2 \geq \|R_1\|_2 - \|R_2\|_2 - \|R_3\|_2 \quad (4)$$

with

$$R_1 = \sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \Big\{ \tanh\left((\theta_a^j)^\top x_{\backslash a}^{(i)}\right), - \tanh\left((\theta_a^{j+1})^\top x_{\backslash a}^{(i)}\right) \Big\}, \quad (5)$$

$$R_2 = \sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \Big\{ \tanh\left((\hat{\theta}_a^{j+1})^\top x_{\backslash a}^{(i)}\right),$$

$$- \tanh\left((\theta_a^{j+1})^\mathsf{T} x_{\backslash a}^{(i)})\right)\Big\}, \quad (6)$$

$$R_3 = \sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} \left\{ x_a^{(i)} - \mathbb{E}_{\Theta^{(j)}}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right]\right\}. \quad (7)$$

The event (4) occurs with probability one and it can be showed that it is included in the event:

$$\{2\lambda_1 + (T_j - \hat{T}_j)\sqrt{p-1}\lambda_2 \geq \frac{1}{3}\|R_1\|_2\}$$
$$\cup \{\|R_2\|_2 \geq \frac{1}{3}\|R_1\|_2\} \cup \{\|R_3\|_2 \geq \frac{1}{3}\|R_1\|_2\}.$$

Thus, we have:

$$\mathbb{P}(A_{n,j} \cap C_n) \leq$$
$$\mathbb{P}(A_{n,j} \cap C_n \cap \{2\lambda_1 + (T_j - \hat{T}_j)\sqrt{p-1}\lambda_2 \geq \frac{1}{3}\|R_1\|_2\})$$
$$+ \mathbb{P}(A_{n,j} \cap C_n \cap \{\|R_2\|_2 \geq \frac{1}{3}\|R_1\|_2\})$$
$$+ \mathbb{P}(A_{n,j} \cap C_n \cap \{\|R_3\|_2 \geq \frac{1}{3}\|R_1\|_2\})$$
$$\triangleq \mathbb{P}(A_{n,j,1}) + \mathbb{P}(A_{n,j,3}) + \mathbb{P}(A_{n,j,3}).$$

Now, We are going to show that each one of the three events has a probability that converges to 0 as $n$ grows. Let's focus on $A_{n,j,1}$. Applying the mean-value theorem, we have for all $i = \widehat{T}_j, ..., T_j - 1$:

$$\tanh\left((\theta_a^j)^\mathsf{T} x_{\backslash a}^{(i)})\right) - \tanh\left((\theta_a^{j+1})^\mathsf{T} x_{\backslash a}^{(i)})\right)$$
$$= (1 - \tanh^2(\bar{\theta}^{iT} x_{\backslash a}^{(i)}))x_{\backslash a}^{(i)\mathsf{T}}(\theta_a^j - \theta_a^{j+1}), \quad (8)$$

with $\bar{\theta}^i = \alpha^i\theta_a^j + (1-\alpha^i)\theta_a^{j+1}$, for a certain $\alpha^i \in [0,1]$. Combining (8) with the definition of $R_1$, we obtain:

$$\|R_1\|_2 = \Bigg\| \sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)}\Big\{ \tanh\left((\theta_a^j)^\mathsf{T} x_{\backslash a}^{(i)})\right)$$
$$- \tanh\left((\theta_a^{j+1})^\mathsf{T} x_{\backslash a}^{(i)})\right)\Big\}\Bigg\|_2 \quad (9)$$

$$= (T_j - \widehat{T}_j)\Bigg\| \frac{1}{T_j - \widehat{T}_j}\sum_{i=\widehat{T}_j}^{T_j-1}(1 - \tanh^2(\bar{\theta}^{iT} x_{\backslash a}^{(i)}))$$
$$\times x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}(\theta_a^j - \theta_a^{j+1})\Bigg\|_2 \quad (10)$$

$$\geq (T_j - \widehat{T}_j)\times$$
$$\times \Lambda_{\min}\left(\frac{1}{T_j - \widehat{T}_j}\sum_{i=\widehat{T}_j}^{T_j-1}(1 - \tanh^2(\bar{\theta}^{iT} x_{\backslash a}^{(i)}))x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}\right)$$
$$\times \|\theta_a^j - \theta_a^{j+1}\|_2 \quad (11)$$

Since, $\forall j,\ \|\theta_a^j\|_2 \leq M$ (A2), we have $\|\bar{\theta}^i\|_2 \leq M$ and $|\bar{\theta}^{iT} x_{\backslash a}^{(i)}| \leq M\cdot\sqrt{p-1}$. Thus, there exist a constant $\tilde{M} > 0$ such that $1 - \tanh^2(\bar{\theta}^{iT} x_{\backslash a}^{(i)}) \geq \tilde{M}$. Combining this with the fact that each matrix $x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}$ are positive semidefinite, we have:

$$\|R_1\|_2 \geq$$
$$(T_j - \widehat{T}_j)\tilde{M}\Lambda_{\min}\left(\frac{1}{T_j - \widehat{T}_j}\sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}\right)\xi_{\min}. \quad (12)$$

Thus, the event $\{2\lambda_1 + (T_j - \hat{T}_j)\sqrt{p-1}\lambda_2 \geq \frac{1}{3}\|R_1\|_2\}$ is included in the event:

$$2\lambda_1 + (T_j - \hat{T}_j)\sqrt{p-1}\lambda_2 \geq$$
$$(T_j - \widehat{T}_j)\tilde{M}\Lambda_{\min}\left(\frac{1}{T_j - \widehat{T}_j}\sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}\right)\xi_{\min}. \quad (13)$$

Denoting by $\{13\}$ the event of Eq. 13, we have:

$$\mathbb{P}(A_{n,j,1}) \leq \mathbb{P}(A_{n,j} \cap C_n \cap \{13\})$$
$$\leq \mathbb{P}\Bigg(A_{n,j} \cap C_n \cap \{13\}$$
$$\cap \{\Lambda_{\min}\left(\frac{1}{T_j - \widehat{T}_j}\sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}\right) > \frac{\phi_{\min}}{2}\}\Bigg)$$
$$+ \mathbb{P}\Bigg(A_{n,j} \cap C_n$$
$$\cap \{\Lambda_{\min}\left(\frac{1}{T_j - \widehat{T}_j}\sum_{i=\widehat{T}_j}^{T_j-1} x_{\backslash a}^{(i)} x_{\backslash a}^{(i)\mathsf{T}}\right) \leq \frac{\phi_{\min}}{2}\}\Bigg).$$

Using Lemma 3 with $v_n = n\delta_n$ and $\epsilon = \frac{\phi_{\min}}{2}$, we can bound the right-hand side of the upper equation. We also re-write the first term so that we obtain:

$$\mathbb{P}(A_{n,j,1})$$
$$\leq \mathbb{P}(A_{n,j} \cap C_n \cap \{\frac{2\lambda_1}{T_j - \widehat{T}_j} + \sqrt{p-1}\lambda_2 > \frac{\tilde{M}\phi_{\min}}{2}\xi_{\min}\})$$
$$+ c_1 \exp\left(-\frac{\epsilon^2 n\delta_n}{2} + 2\log(n)\right)$$
$$\leq \mathbb{P}(\xi_{\min}^{-1}\frac{2\lambda_1}{n\delta_n} + \xi_{\min}^{-1}\sqrt{p-1}\lambda_2 > \frac{\tilde{M}\phi_{\min}}{2})$$
$$+ c_1 \exp\left(-\frac{\epsilon^2 n\delta_n}{2} + 2\log(n)\right).$$

Thanks to $(iii)$, we have $n\delta_n$ that goes to infinity faster than $\log(n)$, thus the second term of the sum goes to 0 as $n$ grows. Furthermore, using $(i)$ and $(ii)$ we have:

$$\mathbb{P}(\xi_{\min}^{-1}\frac{2\lambda_1}{n\delta_n}+\xi_{\min}^{-1}\sqrt{p-1}\lambda_2 > \frac{\tilde{M}\phi_{\min}}{2}) \underset{n\to 0}{\longrightarrow} \mathbb{P}(0+0 > \frac{\tilde{M}\phi_{\min}}{2}) = 0.$$

Which concludes that $\mathbb{P}(A_{n,j,1}) \to 0$.

We now focus on the event $A_{n,j,2}$. Let $\bar{T}_j \triangleq \lfloor 2^{-1}(T_j + T_{j+1})\rfloor$ and remark that between $T_j$ and $\bar{T}_j$, $\widehat{\beta}^{(i)} = \widehat{\theta}^{j+1}$. Now, using Lemma 1 with $k = \bar{T}_j$ and $k = T_j$ and similar operation used to show equation (4), we have:

$$\begin{aligned}
&2\lambda_1 + (\bar{T}_j - T_j)\sqrt{p-1}\lambda_2 \\
&\geq \left\| \sum_{i=T_j}^{\bar{T}_j-1} x_{\backslash a}^{(i)} \Bigg( \tanh\Big((\hat{\theta}_a^{j+1})^\intercal x_{\backslash a}^{(i)}\Big) - \tanh\Big((\theta_a^{j+1})^\intercal x_{\backslash a}^{(i)}\Big)\Bigg)\right\|_2 \\
&- \| \sum_{i=T_j}^{\bar{T}_j-1} x_{\backslash a}^{(i)} \underbrace{\Big(x_a^{(i)} - \mathbb{E}_{\Theta^{(j+1)}}\Big[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\Big]\Big)}_{\varepsilon_{j+1}^i} \|_2.
\end{aligned}$$

Now using the fact that $\|\hat{\theta}_a^{j+1}\|_2$ is necessarily bounded, Lemma 3 with $\epsilon = \phi_{\min}/2$ and similar arguments that we used for $A_{n,j,1}$, we can write that the first term in the right-hand side of the previous equation is lower-bounded by:

$$(T_j - \bar{T}_j)\tilde{\tilde{M}}\frac{\phi_{\min}}{2}\|\hat{\theta}_a^{j+1} - \theta_a^{j+1}\|_2$$

with probability tending to one. Here, $\tilde{\tilde{M}}$ corresponds to a positive constant derived the same way as $\tilde{M}$ in the previous part of the proof. In consequence, we can write

$$\|\hat{\theta}_a^{j+1} - \theta_a^{j+1}\|_2 \leq \frac{8\lambda_1 + 4(\bar{T}_j - T_j)\sqrt{p-1}\lambda_2 + 4\|\sum_{i=T_j}^{\bar{T}_j-1} x_{\backslash a}^{(i)}\varepsilon_{j+1}^i\|_2}{\tilde{\tilde{M}}\phi_{\min}(T_{j+1} - T_j)}, \tag{14}$$

which holds with probability tending to one.

Furthermore, with probability also tending to one it can be shown using the same arguments used to prove equation (12) that $\|R_1\|_2 \geq (T_j - \widehat{T}_j)\tilde{M}\phi_{\min}\xi_{\min}/2$ and $\|R_2\|_2 \leq \|\hat{\theta}_a^{j+1} - \theta_a^{j+1}\|_2\phi_{\max}(T_j - \widehat{T}_j)/2$. Combining that with equation (19), we can write:

$$\begin{aligned}
&\mathbb{P}(A_{n,j,2}) \\
&\leq \mathbb{P}(A_{n,j} \cap C_n \cap \{\frac{1}{3}\tilde{\tilde{M}}\tilde{M}\phi_{\min}^2\phi_{\max}^{-1}\xi_{\min}(T_{j+1} - T_j) \leq \\
&\quad 8\lambda_1 + 4(\bar{T}_j - T_j)\sqrt{p-1}\lambda_2 + 4\|\sum_{i=T_j}^{\bar{T}_j-1} x_{\backslash a}^{(i)}\varepsilon_{j+1}^i\|_2\}) \\
&\quad + c_1\exp\left(-c_2 n\delta_n + 2\log(n)\right) \\
&\leq \mathbb{P}(c_3\phi_{\min}^2\phi_{\max}^{-1}\xi_{\min}\Delta_{\min} \leq \lambda_1) \\
&+ \mathbb{P}(c_4\phi_{\min}^2\phi_{\max}^{-1}\xi_{\min} \leq \sqrt{p-1}\lambda_2) \\
&+ \mathbb{P}\left(c_5\phi_{\min}^2\phi_{\max}^{-1}\xi_{\min} \leq (\bar{T}_j - T_j)^{-1}\|\sum_{i=T_j}^{\bar{T}_j-1} x_{\backslash a}^{(i)}\varepsilon_{j+1}^i\|_2\right) \\
&+ c_1\exp\left(-c_2 n\delta_n + 2\log(n)\right).
\end{aligned}$$

With $c_1, ..., c_5$ positive constants.

The first two terms tends to 0 as $n$ goes to infinity thanks to the hypothesis $(i)$ and $(ii)$ of the theorem. Indeed, since $\Delta_{\min} > n\delta_n$ and $(n\delta_n\xi_{\min})^{-1}\lambda_1 \to 0$ $(i)$, the first term tends to $\mathbb{P}(c_3\phi_{\min}^2\phi_{\max}^{-1} \leq 0) = 0$ and the second term tends to 0 since $\xi_{\min}^{-1}\sqrt{p-1}\lambda_2 \to 0$ $(ii)$. The fourth term directly tends to 0. Applying Lemma 4, we can upper bound the third term by:

$$\begin{aligned}
&\mathbb{P}\left(c_5\phi_{\min}^2\phi_{\max}^{-1}\xi_{\min} \leq (\bar{T}_j - T_j)^{-1/2}2\sqrt{p\log(n)}\right) \\
&\quad + c_6\exp(-2p\log(n)) \\
&\leq \mathbb{P}\left(c_5\phi_{\min}^2\phi_{\max}^{-1}\xi_{\min} \leq (n\delta_n)^{-1/2}2\sqrt{p\log(n)}\right) \\
&\quad + c_6\exp(-2p\log(n))
\end{aligned}$$

with $c_6$ an other positive constant.

Since $(\xi_{\min}\sqrt{n\delta_n})^{-1}\sqrt{p\log(n)} \to 0$ $(iii)$, the previous equation tends to 0, which make $\mathbb{P}(A_{n,j,2})$ tends to 0 as well.

Finally, we upper bound the probability on the event $A_{n,j,3}$. As before, we know that $\|R_1\|_2 \geq (T_j - \widehat{T}_j)\tilde{M}\phi_{\min}\xi_{\min}/2$ with probability at least $1 - c_1\exp(-c_2 n\delta_n + 2\log(n))$, thus we have:

$$\mathbb{P}(A_{n,j,3}) \leq \mathbb{P}\left(\frac{\tilde{M}\phi_{\min}\xi_{\min}}{6} \leq \frac{\|R_3\|_2}{T_j - \widehat{T}_j}\right) + c_1\exp(-c_2 n\delta_n + 2\log(n)).$$

Using Lemma 5, we can upper bound the first term by:

$$\begin{aligned}
&\mathbb{P}\left(\frac{\tilde{M}\phi_{\min}\xi_{\min}}{6} \leq 2\sqrt{\frac{p\log(n)}{T_j - \widehat{T}_j}}\right) + c_2\exp(-c_3\log(n)) \\
&\leq \mathbb{P}\left(\frac{\tilde{M}\phi_{\min}\xi_{\min}}{6} \leq 2\sqrt{\frac{p\log(n)}{n\delta_n}}\right) + c_2\exp(-c_3\log(n)),
\end{aligned}$$

which tends to 0 thanks to $(iii)$. Since the symmetric case follows exactly the same arguments, we have shown that $\mathbb{P}(A_{n,j} \cap C_n) \to 0$. We now need to prove that $\mathbb{P}(A_{n,j} \cap C_n^c) \to 0$.

**Bounding the bad case**

Let us define the following complementary events:

$$D_n^{(l)} \triangleq \left\{ \exists j \in [D], \widehat{T}_j \leq T_{j-1} \right\} \cap C_n^c \quad (15)$$

$$D_n^{(m)} \triangleq \left\{ \forall j \in [D], T_{j-1} < \widehat{T}_j < T_{j+1} \right\} \cap C_n^c \quad (16)$$

$$D_n^{(r)} \triangleq \left\{ \exists j \in [D], \widehat{T}_j \geq T_{j+1} \right\} \cap C_n^c. \quad (17)$$

We can write $\mathbb{P}(A_{n,j} \cap C_n^c) = \mathbb{P}(A_{n,j} \cap D_n^{(l)}) + \mathbb{P}(A_{n,j} \cap D_n^{(m)}) + \mathbb{P}(A_{n,j} \cap D_n^{(r)})$. Again, the goal is to prove that the three terms tends to 0. We will assume that $\widehat{T}_j \leq T_j$ as the other case can be done by symmetry. Let's first focus on the middle term, it has been shown in (Harchaoui and Lévy-Leduc, 2010; Kolar and Xing, 2012; Gibberd and Roy, 2017) that it can be upper bounded in the following way:

$$\begin{aligned}
&\mathbb{P}(A_{n,j} \cap D_n^{(m)}) \\
&\leq \mathbb{P}(A_{n,j} \cap \{(\widehat{T}_{j+1} - T_j) \geq \frac{\Delta_{\min}}{2}\} \cap D_n^{(m)}) \\
&\quad + \mathbb{P}(\{(T_{j+1} - \widehat{T}_{j+1}) \geq \frac{\Delta_{\min}}{2}\} \cap D_n^{(m)}) \\
&\leq \mathbb{P}(A_{n,j} \cap \{(\widehat{T}_{j+1} - T_j) \geq \frac{\Delta_{\min}}{2}\} \cap D_n^{(m)}) \\
&\quad + \sum_{k=j+1}^{D} \mathbb{P}(\{(\widehat{T}_{k+1} - T_k) \geq \frac{\Delta_{\min}}{2}\} \\
&\qquad \cap \{(T_k - \widehat{T}_k) \geq \frac{\Delta_{\min}}{2}\} \cap D_n^{(m)}). \quad (18)
\end{aligned}$$

Let us bound the first term. Assuming the event $A_{n,j} \cap \{(\widehat{T}_{j+1} - T_j) \geq \frac{\Delta_{\min}}{2}\} \cap D_n^{(m)}$ and applying Lemma 1 with $k = \widehat{T}_j$ and $k = T_j$, we can prove similarly as Eq. 19 that:

$$\begin{aligned}
&\|\hat{\theta}_a^{j+1} - \theta_a^j\|_2 \\
&\leq \frac{4\lambda_1 + 2(T_j - \widehat{T}_j)\sqrt{p-1}\lambda_2 + 2\|\sum_{i=\widehat{T}_j}^{T_j-1} x_{\setminus a}^{(i)} \varepsilon_j^i\|_2}{\tilde{\tilde{M}}\phi_{\min}(T_j - \widehat{T}_j)} \\
&\leq c_1\phi_{\min}^{-1}(n\delta_n)^{-1}\lambda_1 + c_2\phi_{\min}^{-1}\sqrt{p-1}\lambda_2 \\
&\qquad + c_3\phi_{\min}^{-1}(T_j - \widehat{T}_j)^{-1}\|\sum_{i=\widehat{T}_j}^{T_j-1} x_{\setminus a}^{(i)}\varepsilon_j^i\|_2
\end{aligned}$$

with probability tending to one. Using Lemma 5 we can bound the third term and obtain:

$$\begin{aligned}
\|\hat{\theta}_a^{j+1} - \theta_a^j\|_2 &\leq c_1\phi_{\min}^{-1}(n\delta_n)^{-1}\lambda_1 + c_2\phi_{\min}^{-1}\sqrt{p-1}\lambda_2 \\
&\quad + c_3\phi_{\min}^{-1}(\sqrt{n\delta_n})^{-1}\sqrt{p\log(n)}
\end{aligned}$$

with probability tending to one. Similarly, applying the same lemmas with $k = T_j$ and either $k = \widehat{T}_{j+1}$, if $\widehat{T}_{j+1} \leq T_{j+1}$ or $k = T_{j+1}$ otherwise, we have:

$$\begin{aligned}
\|\hat{\theta}_a^{j+1} - \theta_a^{j+1}\|_2 &\leq c_4\phi_{\min}^{-1}(n\delta_n)^{-1}\lambda_1 + c_5\phi_{\min}^{-1}\sqrt{p-1}\lambda_2 \\
&\quad + c_6\phi_{\min}^{-1}(\sqrt{n\delta_n})^{-1}\sqrt{p\log(n)}
\end{aligned}$$

with probability tending to one.
Since $\xi_{\min} \leq \|\theta_a^j - \theta_a^{j+1}\|_2 \leq \|\hat{\theta}_a^{j+1} - \theta_a^j\|_2 + \|\hat{\theta}_a^{j+1} - \theta_a^{j+1}\|_2$, we finally upper bound the considered probability by:

$$\begin{aligned}
&\mathbb{P}(A_{n,j} \cap \{(\widehat{T}_{j+1} - T_j) \geq \frac{\Delta_{\min}}{2}\} \cap D_n^{(m)}) \\
&\quad \leq \mathbb{P}(\xi_{\min} \leq c_7\phi_{\min}^{-1}(n\delta_n)^{-1}\lambda_1 + c_8\phi_{\min}^{-1}\sqrt{p-1}\lambda_2 \\
&\quad + c_9\phi_{\min}^{-1}(\sqrt{n\delta_n})^{-1}\sqrt{p\log(n)}).
\end{aligned}$$

this tends to 0 thanks to the hypothesis $(i)$, $(ii)$ and $(iii)$. The other probabilities in the upper bound on $\mathbb{P}(A_{n,j} \cap D_n^{(m)})$ also tends to 0. The proof follows exactly the previous one. We proved that $\mathbb{P}(A_{n,j} \cap D_n^{(m)}) \to 0$, we will now show the same for $\mathbb{P}(A_{n,j} \cap D_n^{(l)})$.

Following (Gibberd and Roy, 2017), we have:

$$\begin{aligned}
\mathbb{P}(D_n^{(l)}) &\leq \sum_{j=1}^{D} 2^{j-1}\mathbb{P}(\max\{l \in [D] : \widehat{T}_l \leq T_{l-1}\}) \\
&\leq 2^{D-1}\sum_{j=1}^{D}\sum_{l>j}\mathbb{P}(\{T_l - \widehat{T}_l \geq \frac{\Delta_{\min}}{2}\} \\
&\qquad \cap \{\widehat{T}_{l+1} - T_l \geq \frac{\Delta_{\min}}{2}\}).
\end{aligned}$$

Now, combining arguments of (Gibberd and Roy, 2017) and those used to bound the elements of (18), we have $\mathbb{P}(D_n^{(l)}) \to 0$. Similarly we can show $\mathbb{P}(D_n^{(r)}) \to 0$ as $n \to 0$. Finally we have $\mathbb{P}(A_{n,j} \cap C_n^c) \to 0$, which concludes the proof. □

**Proposition 1.** *Let $\{x_i\}_{i=1}^n$ be a sequence of observation drawn from the model presented in Sec. 2. Assume the condition of Theorem 1 are respected. Then, if for a fix $D_{\max}$ we have $D \leq \widehat{D} \leq D_{\max}$ then:*

$$\mathbb{P}(d(\widehat{\mathcal{D}}\|\mathcal{D}) \leq n\delta_n) \underset{n\to\infty}{\longrightarrow} 1.$$

*Proof.* Let us show that:

$$\mathbb{P}(\{d(\widehat{\mathcal{D}}\|\mathcal{D}) \geq n\delta_n\} \cap \{D \leq \widehat{D} \leq D_{\max}\})$$

$$\leq \sum_{K=D}^{D_{\max}} \mathbb{P}(\{d(\widehat{\mathcal{D}}\|\mathcal{D}) \geq n\delta_n\} \cap \{\widehat{D} = K\}) \underset{n\to\infty}{\longrightarrow} 0.$$

First, we note that for $K = D$, we have $\mathbb{P}(\{d(\widehat{\mathcal{D}}\|\mathcal{D}) \geq n\delta_n\} \cap \{\widehat{D} = K\}) \underset{n\to\infty}{\longrightarrow} 0$ thanks to Theorem 1. Thus it suffices to show that:

$$\sum_{K=D+1}^{D_{\max}} \mathbb{P}(\{d(\widehat{\mathcal{D}}\|\mathcal{D}) \geq n\delta_n\} \cap \{\widehat{D} = K\})$$
$$\leq \sum_{K=D+1}^{D_{\max}} \sum_{k=1}^{D} \mathbb{P}(\forall 1 \leq l \leq K, |\widehat{T}_l - T_k| \geq n\delta_n) \underset{n\to\infty}{\longrightarrow} 0.$$

Like in (Harchaoui and Lévy-Leduc, 2010), we rewrite the event $\{\forall 1 \leq l \leq K, |\widehat{T}_l - T_k| \geq n\delta_n\}$ as the disjoint union of the events:

$$E_{n,k,1} = \{\forall 1 \leq l \leq K, |\widehat{T}_l - T_k| \geq n\delta_n \text{ and } \widehat{T}_l < T_k\}$$
$$E_{n,k,2} = \{\forall 1 \leq l \leq K, |\widehat{T}_l - T_k| \geq n\delta_n \text{ and } \widehat{T}_l > T_k\}$$
$$E_{n,k,3} = \{\exists 1 \leq l \leq K-1, |\widehat{T}_l - T_k| \geq n\delta_n, |\widehat{T}_{l+1} - T_k| \geq n\delta_n \text{ and } \widehat{T}_l < T_k < \widehat{T}_{l+1}\}$$

and propose to show that the probability of each events tends to 0 as $n$ grows. Let's begin with $\mathbb{P}(E_{n,k,1})$ and note that it is equal to:

$$\mathbb{P}(E_{n,k,1} \cap \{\widehat{T}_K > T_{k-1}\}) + \mathbb{P}(E_{n,k,1} \cap \{\widehat{T}_K \leq T_{k-1}\})$$

First, we are going to upper bound the left-hand element of the previous equation. Applying Lemma 1 with $t = \widehat{T}_K$ and $t = T_k$, we can prove similarly to the equation (4) in the good case scenario of the previous theorem that:

$$2\lambda_1 + (T_k - \widehat{T}_K)\sqrt{p-1}\lambda_2 \geq \|R_1'\|_2 - \|R_2'\|_2 - \|R_3'\|_2$$

with

$$R_1' = \sum_{i=\widehat{T}_K}^{T_k-1} x_{\backslash a}^{(i)} \Big\{ \tanh\Big((\theta_a^k)^T x_{\backslash a}^{(i)}\Big) - \tanh\Big((\theta_a^{k+1})^T x_{\backslash a}^{(i)}\Big) \Big\}$$
$$R_2' = \sum_{i=\widehat{T}_K}^{T_k-1} x_{\backslash a}^{(i)} \Big\{ \tanh\Big((\hat{\theta}_a^{K+1})^T x_{\backslash a}^{(i)}\Big) - \tanh\Big((\theta_a^{k+1})^T x_{\backslash a}^{(i)}\Big) \Big\}$$
$$R_3' = \sum_{i=\widehat{T}_K}^{T_k-1} x_{\backslash a}^{(i)} \Big\{ x_a^{(i)} - \mathbb{E}_{\Theta^{(k)}}\Big[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\Big] \Big\}.$$

Like in the previous theorem, we can upperbound $\mathbb{P}(E_{n,k,1} \cap \{\widehat{T}_k > T_{k-1}\})$ by:

$$\mathbb{P}(E_{n,k,1}^{(1)}) + \mathbb{P}(E_{n,k,1}^{(2)}) + \mathbb{P}(E_{n,k,1}^{(3)})$$

where

$$E_{n,k,1}^{(1)} = \{2\lambda_1 + (T_k - \widehat{T}_K)\sqrt{p-1}\lambda_2 \geq \frac{1}{3}\|R_1'\|_2\}$$
$$E_{n,k,1}^{(2)} = \{\|R_2'\|_2 \geq \frac{1}{3}\|R_1'\|_2\}$$
$$E_{n,k,1}^{(3)} = \{\|R_3'\|_2 \geq \frac{1}{3}\|R_1'\|_2\}.$$

To show that $\mathbb{P}(E_{n,k,1}^{(1)})$ tends to 0 it suffices to follow the proof used to show that $\mathbb{P}(A_{n,j,1})$ tends to 0 in the good scenario of the previous theorem.

Similarly, to show that $\mathbb{P}(E_{n,k,1}^{(2)})$ tends to 0 it suffices to follow the proof used for $\mathbb{P}(A_{n,j,2})$. Applying lemma 1 with $t = T_k$ ans $t = T_{k+1}$ we can show that with probability tending to one:

$$\|\widehat{\theta}_a^{K+1} - \theta_a^{k+1}\|_2 \leq \frac{4\lambda_1 + 2(T_{k+1} - T_k)\sqrt{p-1}\lambda_2 + 2\|\sum_{i=T_k}^{T_{k+1}} x_{\backslash a}^{(i)} \varepsilon_{j+1}^i\|_2}{\tilde{\tilde{M}}\phi_{\min}(T_{k+1} - T_k)}. \tag{19}$$

The rest follows exactly the arguments used to show the limit of $\mathbb{P}(A_{n,j,2})$.

Finally, $\mathbb{P}(E_{n,k,1}^{(3)})$ tends to 0 the same way $\mathbb{P}(A_{n,j,3})$ was tending to 0 in the previous proof.

The proof to show that $\mathbb{P}(E_{n,k,1} \cap \{\widehat{T}_K \leq T_{k-1}\})$ tends to 0 is the same. It suffices to apply lemma 1 with $t = T_{k-1}$ and $t = T_k$ to split the event in 3 sub-events and follow the proof. By symmetry, we also have $\mathbb{P}(E_{n,k,2}) \to 0$.

Let's now focus on $E_{n,k,3}$. Like in (Harchaoui and Lévy-Leduc, 2010), the event is split is four independent events:

$$E_{n,k,3} = E_{n,k,3}^{(1)} \cup E_{n,k,3}^{(2)} \cup E_{n,k,3}^{(3)} \cup E_{n,k,3}^{(4)}$$

with

$$E_{n,k,3}^{(1)} = E_{n,k,3} \cap \{T_{k-1} < \widehat{T}_l < \widehat{T}_{l+1} < T_{k+1}\}$$
$$E_{n,k,3}^{(2)} = E_{n,k,3} \cap \{T_{k-1} < \widehat{T}_l < T_{k+1}, \widehat{T}_{l+1} > T_{k+1}\}$$
$$E_{n,k,3}^{(3)} = E_{n,k,3} \cap \{\widehat{T}_l < T_{k-1}, T_{k-1} < \widehat{T}_{l+1} < T_{k+1}\}$$
$$E_{n,k,3}^{(4)} = E_{n,k,3} \cap \{\widehat{T}_l < T_{k-1}, \widehat{T}_{l+1} > T_{k+1}\}.$$

To prove that each one of the previous events have a probability that tends to 0 as $n$ grows, we invite the reader to read the proof of (Harchaoui and Lévy-Leduc, 2010). It consist in multiple applications of the different Lemmas, the same way we used them in the previous part. Only the time at which lemma 1 is used changes and are given by (Harchaoui and Lévy-Leduc, 2010). This concludes the proof. □

## Supplementary Lemmas

Below, the different lemmas necessary to prove the main results are given.

**Lemma 2.** *Let $\{x^{(i)}\}_{i=1}^n$ be a set of i.i.d observation sampled from an Ising model with parameter $\Theta \in \mathbb{R}^{p\times p}$ and assume that assumption (A1) is satisfied. Then, $\forall r, l \in [n]$ such that $l < r$ and $r - l > v_n$ with $v_n$ a positive serie, we have $\forall \epsilon > 0$:*

$$\mathbb{P}\left(\Lambda_{\min}\left(\frac{1}{r-l+1}\sum_{i=l}^{r} x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}}\right) \leq \phi_{\min} - \epsilon\right) \leq 2(p-1)^2 \exp\left(-\frac{\epsilon^2 v_n}{2}\right) \quad (20)$$

*and*

$$\mathbb{P}\left(\Lambda_{\max}\left(\frac{1}{r-l+1}\sum_{i=l}^{r} x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}}\right) \geq \phi_{\max} + \epsilon\right) \leq 2(p-1)^2 \exp\left(-\frac{\epsilon^2 v_n}{2}\right). \quad (21)$$

*Proof.* Let $\widehat{\Sigma} = \frac{1}{r-l+1}\sum_{i=l}^{r} x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}}$ and $\Sigma = \mathbb{E}\left[X_{\setminus a} X_{\setminus a}^{\mathsf{T}}\right]$.

We first proove the inequality (20). Recall that for a symmetric matrix $M$, we have $\Lambda_{\max}(M) \leq \|M\|_F$, the Frobenius norm of $M$. We have

$$\begin{aligned}
\Lambda_{\min}(\widehat{\Sigma}) &= \min_{\|v\|_2=1} v^{\mathsf{T}}\widehat{\Sigma}v && (22)\\
&\geq \min_{\|v\|_2=1} v^{\mathsf{T}}\Sigma v - \max_{\|v\|_2=1} v^{\mathsf{T}}(\widehat{\Sigma}-\Sigma)v && (23)\\
&\geq \Lambda_{\min}(\Sigma) - \Lambda_{\max}(\widehat{\Sigma}-\Sigma) && (24)\\
&\geq \phi_{\min} - \|\widehat{\Sigma}-\Sigma\|_F. && (25)
\end{aligned}$$

Let $s_{mq}^{(i)}$ be the $(m,q)$-th coordinate of $x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}} - \Sigma$ and $\frac{1}{r-l+1}\sum_{i=l}^{r} s_{mq}^{(i)}$ the one of $\widehat{\Sigma} - \Sigma$. Note that $\mathbb{E}\left[s_{mq}^{(i)}\right] = 0$ and $|s_{mq}^{(i)}| \leq 2$. Let us analyze the quantity $\mathbb{P}\left(\|\widehat{\Sigma}-\Sigma\|_F > \epsilon\right)$ with $\epsilon > 0$:

$$\begin{aligned}
\mathbb{P}\left(\|\widehat{\Sigma}-\Sigma\|_F > \epsilon\right) &= \mathbb{P}\left((\sum_{m,q} s_{mq}^2)^{1/2} > \epsilon\right) && (26)\\
&= \mathbb{P}\left(\sum_{m,q} s_{mq}^2 > \epsilon^2\right) && (27)\\
&\leq \sum_{m,q} \mathbb{P}\left(s_{mq}^2 > \epsilon^2\right) && (28)\\
&\leq \sum_{m,q} \mathbb{P}\left(|s_{mq}| > \epsilon\right). && (29)
\end{aligned}$$

Thanks to Hoeffding's inequality, we have $\mathbb{P}\left(|s_{mq}| > \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2(r-l+1)}{2}\right)$. Since $r - l > v_n$, we also have $\mathbb{P}\left(|s_{mq}| > \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2 v_n}{2}\right)$. It follows from (29) that $\mathbb{P}\left(\|\widehat{\Sigma}-\Sigma\|_F > \epsilon\right) \leq 2(p-1)^2\exp\left(-\frac{\epsilon^2 v_n}{2}\right)$. We deduce that:

$$\mathbb{P}\left(\Lambda_{\min}(\widehat{\Sigma}) \geq \phi_{\min} - \epsilon\right) \geq 1 - 2(p-1)^2\exp\left(-\frac{\epsilon^2 v_n}{2}\right), \quad (30)$$

which concludes the proof for (20).

To prove (21) it suffices to note that $\Lambda_{\max}(\widehat{\Sigma}) \leq \phi_{\max} + \|\widehat{\Sigma}-\Sigma\|_F$ and use the same arguments. □

**Lemma 3.** *Let $\{x^{(i)}\}_{i=1}^n$ be a set of i.i.d observation sampled from an Ising model with parameter $\Theta \in \mathbb{R}^{p\times p}$ and assume that assumption (A1) is satisfied.*
*Let $R$ and $L$ be two random variable such that $R, L \in [n]$, $L < R$ and $R - L > v_n$ almost surely, with $v_n$ a positive serie. For a fixed node $a$ and any $\epsilon > 0$, there exist a constant $c_1 > 0$ such that:*

$$\mathbb{P}\left(\Lambda_{\min}\left(\frac{1}{R-L+1}\sum_{i=L}^{R} x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}}\right) \leq \phi_{\min} - \epsilon\right) \leq c_1 \exp\left(-\frac{\epsilon^2 v_n}{2} + 2\log(n)\right) \quad (31)$$

*and*

$$\mathbb{P}\left(\Lambda_{\max}\left(\frac{1}{R-L+1}\sum_{i=L}^{R} x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}}\right) \geq \phi_{\max} + \epsilon\right) \leq c_1 \exp\left(-\frac{\epsilon^2 v_n}{2} + 2\log(n)\right). \quad (32)$$

*Proof.* We note $\widehat{\Sigma}(L,R) = \frac{1}{R-L+1}\sum_{i=L}^{R} x_{\setminus a}^{(i)} x_{\setminus a}^{(i)\mathsf{T}}$ and $\mathcal{I} \triangleq \left\{(l,r) \in [n]^2 : r - l > v_n\right\}$.
We first prove the inequality (31):

$$\begin{aligned}
&\mathbb{P}\left(\Lambda_{\max}\left(\widehat{\Sigma}(L,R)\right) \geq \phi_{\max} + \epsilon\right) && (33)\\
&= \sum_{(l,r)\in\mathcal{I}} \mathbb{P}\left(\Lambda_{\max}\left(\widehat{\Sigma}(L,R)\right), L = l, R = r\right) && (34)\\
&\leq \sum_{(l,r)\in\mathcal{I}} \mathbb{P}\left(\Lambda_{\max}\left(\widehat{\Sigma}(L,R)\right)\Big| L = l, R = r\right). && (35)
\end{aligned}$$

Using Lemma 2 we can bound (35):

$$(35) \leq \sum_{(l,r)\in\mathcal{I}} 2(p-1)^2 \exp\left(-\frac{\epsilon^2 v_n}{2}\right) \quad (36)$$

$$\leq |\mathcal{I}| c_1 \exp\left(-\frac{\epsilon^2 v_n}{2}\right) \tag{37}$$

$$\leq n^2 c_1 \exp\left(-\frac{\epsilon^2 v_n}{2}\right) \tag{38}$$

$$\leq c_1 \exp\left(-\frac{\epsilon^2 v_n}{2} + 2\log(n)\right) \tag{39}$$

with $c_1 = 2(p-1)$. This concludes the proof for (31). Same arguments are used to prove (32). □

**Lemma 4.** *Let $\{x^{(i)}\}_{i=1}^n$ be a set of independent observation sampled from the time-varying Ising model (Section 2). Then, $\forall j \in [D]$ and $\forall r, l \in \{T_j, ..., T_{j+1}-1\}$ such that $l < r$, we have:*

$$\mathbb{P}\left(\frac{1}{r-l+1}\|R_3(l,r)\|_2 \leq 2\sqrt{\frac{p\log(n)}{r-l+1}}\right) \tag{40}$$

$$\geq 1 - 2(p-1)\exp\left(-2p\log(n)\right) \tag{41}$$

*with* $R_3(l,r) = \sum_{i=l}^{r} x_{\backslash a}^{(i)} \left\{x_a^{(i)} - \mathbb{E}_{\Theta^j}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right]\right\}$.

*Proof.* Let $Z_{ij}$ be the the $j$-th element of the vector $\frac{1}{r-l+1} x_{\backslash a}^{(i)} \left\{x_a^{(i)} - \mathbb{E}_{\Theta}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right]\right\}$. Note that $|Z_{ij}| \leq \frac{2}{r-l+1}$ and $\mathbb{E}\left[Z_{ij}\right] = 0$. Let $\epsilon > 0$, we have:

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{r-l+1}\|R_3(l,r)\|_2 \geq \epsilon\right) \\
&= \mathbb{P}\left(\sqrt{\sum_{j\neq a}(\sum_{i=l}^{r} Z_{ij})^2} \geq \epsilon\right) \\
&= \mathbb{P}\left(\sum_{j\neq a}(\sum_{i=l}^{r} Z_{ij})^2 \geq \epsilon^2\right) \\
&\leq \sum_{j\neq a} \mathbb{P}\left(|\sum_{i=l}^{r} Z_{ij}| \geq \epsilon\right) \\
&\leq 2(p-1)\exp\left(-\frac{\epsilon^2(r-l+1)}{2}\right).
\end{aligned}$$

Now, if we fix $\epsilon = 2\sqrt{\frac{p\log(n)}{r-l+1}}$, we obtain:

$$\mathbb{P}\left(\frac{1}{r-l+1}\|R_3(l,r)\|_2 \leq 2\sqrt{\frac{p\log(n)}{r-l+1}}\right) \geq 1 - 2(p-1)\exp\left(-2p\log(n)\right).$$

□

**Lemma 5.** *Let $\{x^{(i)}\}_{i=1}^n$ be a set of independent observation sampled from the time-varying Ising model (Section 2). We have:*

$$\mathbb{P}\left(\bigcap_{j\in[D]}\bigcap_{l,r\in\mathcal{I}_j}\left\{\frac{1}{r-l+1}\|R_3^j(l,r)\|_2 \leq 2\sqrt{\frac{p\log(n)}{r-l+1}}\right\}\right) \geq 1 - c_2\exp\left(-c_3\log(n)\right) \tag{42}$$

*with* $R_3^j(l,r) = \sum_{i=l}^{r} x_{\backslash a}^{(i)} \left\{x_a^{(i)} - \mathbb{E}_{\Theta^j}\left[X_a | X_{\backslash a} = x_{\backslash a}^{(i)}\right]\right\}$, $c_2, c_3$ *some positive constants and* $\mathcal{I}_j \triangleq \left\{(l,r) \in \{T_j, ..., T_{j+1}-1\}^2 : r > l\right\}$.

*Proof.* The proof is a simple application of Lemma 4:

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{j\in[D]}\bigcup_{l,r\in\mathcal{I}_j}\left\{\frac{1}{r-l+1}\|R_3^j(l,r)\|_2 \geq 2\sqrt{\frac{p\log(n)}{r-l+1}}\right\}\right) \\
&\leq \sum_{j\in[D]}\sum_{l,r\in\mathcal{I}_j} \mathbb{P}\left(\frac{1}{r-l+1}\|R_3^j(l,r)\|_2 \geq 2\sqrt{\frac{p\log(n)}{r-l+1}}\right) \\
&\leq 2Dn^2(p-1)\exp\left(-2p\log(n)\right) \\
&\leq c_2\exp\left(-2p\log(n) + 2\log(n)\right) \\
&\leq c_2\exp\left(-c_3\log(n)\right)
\end{aligned}$$

since $p > 1$. This concludes the proof. □

## References


Gibberd, A. J. and Roy, S. (2017). Multiple changepoint estimation in high-dimensional gaussian graphical models. *arXiv preprint arXiv:1712.05786*.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492):1480–1493.

Kolar, M. and Xing, E. P. (2012). Estimating networks with jumps. *Electronic journal of statistics*, 6:2069.

Le Bars, B. and Kalogeratos, A. (2019). A probabilistic framework to node-level anomaly detection in communication networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 2188–2196. IEEE.