Supplement to the paper “Change-Point Estimation in High-Dimensional Markov Random Field Models”

Sandipan Roy†, Yves Atchadé† and George Michailidis†

University of Michigan, Ann Arbor, USA.

Although our main motivation is in discrete graphical models, the proposed methodology can be applied more broadly for model-based change-point estimation. With this in mind, we shall prove a more general result that can be useful with other high-dimensional change-point estimation problems. Theorem 1 follows as a special case.

S1. High-dimensional model-based change-point detection

Let $\{X^{(t)},\ 1 \le t \le T\}$ be a sequence of $\mathbb{R}^p$-valued independent random variables. Let $\Theta \subseteq \mathbb{R}^d$ be an open, non-empty convex parameter space equipped with the Euclidean inner product $\langle \cdot, \cdot \rangle$, and norm$\|\cdot\|_2$. We will also use the ℓ^1-norm $\|\theta\|_1 \stackrel{\text{def}}{=} \sum_{j=1}^d |\theta_j|$, and the ℓ^∞-norm $\|\theta\|_\infty \stackrel{\text{def}}{=} \max_{1\le j\le d} |\theta_j|$. We assume that there exists a change point $\tau_\star \in \{1, \ldots, T-1\}$, parameters $\theta_\star^{(1)}, \theta_\star^{(2)} \in \Theta$, such that for $t = 1, \ldots, \tau_\star$, $X^{(t)} \sim g^{(t)}_{\theta_\star^{(1)}}$, and for $t = \tau_\star + 1, \ldots, T$, $X^{(t)} \sim g^{(t)}_{\theta_\star^{(2)}}$, where $g^{(t)}_{\theta_\star^{(1)}}$ and $g^{(t)}_{\theta_\star^{(2)}}$ are probability densities on $\mathbb{R}^p$. The goal is to estimate $\tau_\star, \theta_\star^{(1)}, \theta_\star^{(2)}$. This setting includes the Markov random field setting (our main motivation), where $g^{(t)}_{\theta_\star^{(1)}}$ and $g^{(t)}_{\theta_\star^{(2)}}$ does not depend t. It also includes regression models where the index t in the distributions $g^{(t)}_{\theta_\star^{(1)}}$ and $g^{(t)}_{\theta_\star^{(2)}}$ accounts for the covariates of subject t.

For $t = 1 \ldots, T$, let $(\theta, x) \mapsto \phi_t(\theta, x)$ be jointly measurable functions on $\Theta \times \mathbb{R}^p$, such that $\theta \mapsto \phi_t(\theta, x)$ is convex and continuously differentiable for all $x \in \mathbb{R}^p$. We define

$$\ell_T(\tau; \theta_1, \theta_2) \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{\tau} \phi_t(\theta_1, X^{(t)}) + \frac{1}{T}\sum_{t=\tau+1}^{T} \phi_t(\theta_2, X^{(t)}),$$

†*Address for correspondence:* Department of Statistics, 439 West Hall, University of Michigan, Ann Arbor, MI 48109-1107, USA

E-mail: sandipan@umich.edu, yvesa@umich.edu, gmichail@umich.edu

and we consider the change-point estimator $\tau_\star$ given by

$$\widehat{\tau} = \operatorname*{Argmin}_{\tau \in \mathcal{T}} \ell_T(\tau; \widehat{\theta}_{1,\tau}, \widehat{\theta}_{2,\tau}), \tag{S1}$$

for a non-empty search domain $\mathcal{T} \subset \{1, \ldots, T\}$, where for each $\tau \in \mathcal{T}$, $\widehat{\theta}_{1,\tau}$ and $\widehat{\theta}_{2,\tau}$ are defined as

$$\widehat{\theta}_{1,\tau} \stackrel{\text{def}}{=} \operatorname*{Argmin}_{\theta \in \Theta} \left[\frac{1}{T} \sum_{t=1}^{\tau} \phi_t(\theta, X^{(t)}) + \lambda_{1,\tau} \|\theta\|_1 \right],$$

and

$$\widehat{\theta}_{2,\tau} \stackrel{\text{def}}{=} \operatorname*{Argmin}_{\theta \in \Theta} \left[\frac{1}{T} \sum_{t=\tau+1}^{T} \phi_t(\theta, X^{(t)}) + \lambda_{2,\tau} \|\theta\|_1 \right],$$

for some positive penalty parameters $\lambda_{1,\tau}, \lambda_{2,\tau}$. Note that by allowing the use of user-defined learning functions ϕ_t, our framework can be used to analyze maximum likelihood and maximum pseudo-likelihood change-point estimators.

For $\tau \in \{1, \ldots, T-1\}$, we set

$$\mathcal{G}^1_\tau \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=1}^{\tau} \nabla \phi_t(\theta_\star^{(1)}, X^{(t)}), \quad \text{and} \quad \mathcal{G}^2_\tau \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=\tau+1}^{T} \nabla \phi_t(\theta_\star^{(2)}, X^{(t)}),$$

where $\nabla \phi_t(\theta, x)$ denotes the partial derivative of $u \mapsto \phi_t(u, x)$ at θ. Also for $\tau \in \{1, \ldots, T-1\}$, and for $\theta \in \Theta$, we define,

$$\mathcal{L}_1(\tau, \theta) \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=1}^{\tau} \left[\phi_t(\theta, X^{(t)}) - \phi_t(\theta_\star^{(1)}, X^{(t)}) - \left\langle \nabla \phi_t(\theta_\star^{(1)}, X^{(t)}), \theta - \theta_\star^{(1)} \right\rangle \right],$$

$$\text{and} \quad \mathcal{L}_2(\tau, \theta) \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=\tau+1}^{T} \left[\phi_t(\theta, X^{(t)}) - \phi_t(\theta_\star^{(2)}, X^{(t)}) - \left\langle \nabla \phi_t(\theta_\star^{(2)}, X^{(t)}), \theta - \theta_\star^{(2)} \right\rangle \right].$$

For $j = 1, 2$, define $\mathcal{A}_j \stackrel{\text{def}}{=} \left\{ 1 \leq k \leq d : \theta_{\star k}^{(j)} \neq 0 \right\}$, $s_j = |\mathcal{A}_j|$, and

$$\mathbb{C}_j \stackrel{\text{def}}{=} \left\{ \theta \in \Theta : \sum_{k \in \mathcal{A}_j^c} |\theta_k^{(j)}| \leq 3 \sum_{k \in \mathcal{A}_j} |\theta_k^{(j)}| \right\}. \tag{S2}$$

The curvature of the function $\mathcal{L}_j(\tau, \cdot)$ is not always best described with the usual quadratic function $\theta \mapsto \|\theta - \theta_\star^{(j)}\|_2^2$. We will need a more flexible framework, in order to handle $\mathcal{L}_j(\tau, \cdot)$ in the case of discrete Markov random fields. Let $\mathsf{r} : [0, \infty) \to [0, \infty)$ be continuous function such that $x \mapsto \mathsf{r}(x)/x$ is strictly increasing and $\lim_{x\downarrow 0} \mathsf{r}(x)/x = 0$. We call r a rate function, and for $a > 0$, we define $\Psi_{\mathsf{r}}(a) \stackrel{\text{def}}{=} \inf\{x > 0 :\ \mathsf{r}(x)/x \geq a\}$ ($\inf \emptyset = +\infty$). For $\tau \in \{1, \ldots, T-1\}$, $\lambda > 0$, a rate function r, $c > 0$, and for $j = 1, 2$ we work with the event

$$\mathcal{E}_\tau^j(\lambda, \mathsf{r}, c) \stackrel{\text{def}}{=} \left\{ \|G_\tau^j\|_\infty \leq \frac{\lambda}{2},\quad \inf_{\theta \neq \theta_\star^{(j)},\ \theta - \theta_\star^{(j)} \in \mathbb{C}_j} \frac{\mathcal{L}_j(\tau, \theta)}{\mathsf{r}\left(\|\theta - \theta_\star^{(j)}\|_2\right)} \geq \frac{\tau}{T}, \right.$$
$$\left. \sup_{\theta \neq \theta_\star^{(j)},\ \theta - \theta_\star^{(j)} \in \mathbb{C}_j} \frac{\mathcal{L}_j(\tau, \theta)}{\|\theta - \theta_\star^{(j)}\|_2^2} \leq \frac{\tau}{T} \frac{c}{2} \right\}.$$

Define

$$\kappa_0^{(t)} \stackrel{\text{def}}{=} \begin{cases} \mathbb{E}\left[\phi_t(\theta_\star^{(2)}, X^{(t)}) - \phi_t(\theta_\star^{(1)}, X^{(t)})\right] & \text{if } t \leq \tau_\star \\ \mathbb{E}\left[\phi_t(\theta_\star^{(1)}, X^{(t)}) - \phi_t(\theta_\star^{(2)}, X^{(t)})\right] & \text{if } t > \tau_\star \end{cases},$$

and

$$U^{(t)} \stackrel{\text{def}}{=} \begin{cases} \phi_t(\theta_\star^{(2)}, X^{(t)}) - \phi_t(\theta_\star^{(1)}, X^{(t)}) - \kappa_0^{(t)} & \text{if } t \leq \tau_\star \\ \phi_t(\theta_\star^{(1)}, X^{(t)}) - \phi_t(\theta_\star^{(2)}, X^{(t)}) - \kappa_0^{(t)} & \text{if } t > \tau_\star \end{cases}.$$

We make the following assumption.

A 1. *There exist finite constants $\sigma_{0t} > 0$ such that*

$$\mathbb{E}\left(e^{xU^{(t)}}\right) \leq e^{x^2 \sigma_{0t}^2 \|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2/2}, \quad \textit{for all} \quad x > 0.$$

Furthermore, there exist $B_0 > 0$, $\bar{\sigma}_0^2 > 0$, $\bar{\kappa}_0 > 0$ such that for all integer $k \geq B_0$,

$$\min\left(\frac{1}{k} \sum_{t=\tau_\star-k+1}^{\tau_\star} \kappa_0^{(t)},\ \frac{1}{k} \sum_{t=\tau_\star+1}^{\tau_\star+k} \kappa_0^{(t)}\right) \geq \bar{\kappa}_0 \|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2, \tag{S3}$$

and

$$\max\left(\frac{1}{k} \sum_{t=\tau_\star-k+1}^{\tau_\star} \sigma_{0t}^2, \frac{1}{k} \sum_{t=\tau_\star+1}^{\tau_\star+k} \sigma_{0t}^2\right) \leq \bar{\sigma}_0^2. \tag{S4}$$

THEOREM S1. *Assume A1, and $\theta_\star^{(1)} \neq \theta_\star^{(2)}$. Suppose that $\hat{\tau}$ is defined over a search domain $\mathcal{T} \ni \tau_\star$, and with penalty $\lambda_{j,\tau} > 0$ (for $j = 1, 2$). For $j = 1, 2$, take a rate function r_j, constant $c_j > 0$, and define $\mathcal{E} \stackrel{\text{def}}{=} \cap_{\tau \in \mathcal{T}} \mathcal{E}_\tau^1(\lambda_{1,\tau}, \mathsf{r}_1, c_1) \cap \mathcal{E}_\tau^2(\lambda_{2,\tau}, \mathsf{r}_2, c_2)$. Set*

$$\delta(\tau) \stackrel{\text{def}}{=} \Psi_{r_1}\left(6\left(\frac{T}{\tau}\right) s_1^{1/2} \lambda_{1,\tau}\right) \left[2 s_1^{1/2} T \lambda_{1,\tau} + \tau \Psi_{r_1}\left(6\left(\frac{T}{\tau}\right) s_1^{1/2} \lambda_{1,\tau}\right)\right]$$
$$+ \Psi_{r_2}\left(6\left(\frac{T}{T-\tau}\right) s_2^{1/2} \lambda_{2,\tau}\right) \left[2 s_2^{1/2} T \lambda_{2,\tau} + (T-\tau) \Psi_{r_2}\left(6\left(\frac{T}{T-\tau}\right) s_2^{1/2} \lambda_{2,\tau}\right)\right],$$

$\delta \stackrel{\text{def}}{=} \sup_{\tau\in\mathcal{T}} \delta(\tau)$, *and* $B \stackrel{\text{def}}{=} \max\left(B_0, \frac{4\delta}{\bar{\kappa}_0\|\theta_\star^{(2)}-\theta_\star^{(1)}\|_2^2}\right)$, *with* B_0 *as in A1. Then*

$$\mathbb{P}\left(|\hat{\tau}-\tau_\star| > B\right) \le 2\mathbb{P}(\mathcal{E}^c) + \frac{4\exp\left(-\frac{\bar{\kappa}_0^2\delta}{2\bar{\sigma}_0^2}\right)}{1-\exp\left(-\frac{\bar{\kappa}_0^2\|\theta_\star^{(2)}-\theta_\star^{(1)}\|_2^2}{8\bar{\sigma}_0^2}\right)}. \tag{S5}$$

PROOF. The starting point of the proof is the following variant of a result due to Neghaban et al. (2010).

LEMMA 1. *Fix* $\tau \in \{1,2,\ldots,T-1\}$. *On* $\mathcal{E}_\tau^1(\lambda_{1,\tau}, \mathsf{r}_1, c_1) \cap \mathcal{E}_\tau^2(\lambda_{2,\tau}, \mathsf{r}_2, c_2)$, $\hat{\theta}_{j,\tau} - \theta_\star^{(j)} \in \mathbb{C}_j$, $(j=1,2)$, *where* $\mathbb{C}_j$ *is defined in (S2), and*

$$\|\hat{\theta}_{1,\tau} - \theta_\star^{(1)}\|_2 \le \Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right) s_1^{1/2}\lambda_{1,\tau}\right),$$

$$\text{and} \quad \|\hat{\theta}_{2,\tau} - \theta_\star^{(2)}\|_2 \le \Psi_{\mathsf{r}_2}\left(6\left(\frac{T}{T-\tau}\right) s_2^{1/2}\lambda_{2,\tau}\right). \tag{S6}$$

PROOF. We prove the first inequality. The second follows similarly. We set

$$\mathcal{U}(\theta) \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{\tau}\phi_t(\theta, X^{(t)}) + \lambda_{1,\tau}\|\theta\|_1 - \left(\frac{1}{T}\sum_{t=1}^{\tau}\phi_t(\theta_\star^{(1)}, X^{(t)}) + \lambda_{1,\tau}\|\theta_\star^{(1)}\|_1\right).$$

Since $\hat{\theta}_{1,\tau} = \mathsf{Argmin}_{\theta\in\Theta}\left[\frac{1}{T}\sum_{t=1}^{\tau}\phi_t(\theta, X^{(t)}) + \lambda_{1,\tau}\|\theta\|_1\right]$, and using the convexity of the functions ϕ_t we have

$$0 \ge \mathcal{U}(\hat{\theta}_{1,\tau}) \ge \left\langle G_\tau^1, \hat{\theta}_{1,\tau} - \theta_\star^{(1)}\right\rangle + \lambda_{1,\tau}\left(\|\hat{\theta}_{1,\tau}\|_1 - \|\theta_\star^{(1)}\|_1\right).$$

On $\mathcal{E}_\tau^1(\lambda_{1,\tau}, \mathsf{r}_1, c_1)$, $\|G_\tau^1\|_\infty \le \lambda_{1,\tau}/2$. Using this and some easy algebra as in Neghaban et al. (2010), shows that $\hat{\theta}_{1,\tau} - \theta_\star^{(1)} \in \mathbb{C}_1$. Set $b = \Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)$. We will show that for all $\theta \in \mathbb{R}^d$ such that $\theta - \theta_\star^{(1)} \in \mathbb{C}_1$, and $\|\theta - \theta_\star^{(1)}\|_2 > b$, we have $\mathcal{U}(\theta) > 0$. Since $\mathcal{U}(\hat{\theta}_{1,\tau}) \le 0$, and $\hat{\theta}_{1,\tau} - \theta_\star^{(1)} \in \mathbb{C}_1$, the claim that $\|\theta - \theta_\star^{(1)}\|_2 \le b$ follows. On the event $\mathcal{E}_\tau^1(\lambda_{1,\tau}, \mathsf{r}_1, c_1)$, and for $\theta - \theta_\star^{(1)} \in \mathbb{C}_1$, we have

$$\begin{aligned}
\mathcal{U}(\theta) &= \left\langle G_\tau^1, \theta - \theta_\star^{(1)}\right\rangle + \mathcal{L}_1(\tau,\theta) + \lambda_{1,\tau}\left(\|\theta\|_1 - \|\theta_\star^{(1)}\|_1\right) \\
&\ge \frac{\tau}{T}\mathsf{r}_1(\|\theta - \theta_\star^{(1)}\|_2) - \frac{3\lambda_{1,\tau}}{2}\|\theta - \theta_\star^{(1)}\|_1 \\
&\ge \frac{\tau}{T}\left[\mathsf{r}_1(\|\theta - \theta_\star^{(1)}\|_2) - 6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\|\theta - \theta_\star^{(1)}\|_2\right].
\end{aligned}$$

Using the definition of Ψ_{r_1}, we then see that $\mathcal{U}(\theta) > 0$ for $\|\theta - \theta_\star^{(1)}\|_2 > b$. This ends the proof. $\square$

The next result follows easily.

LEMMA 2. *Fix* $\tau \in \{1, 2, \ldots, T-1\}$. *On* $\mathcal{E}_\tau^1(\lambda_{1,\tau}, \mathsf{r}_1, c_1) \cap \mathcal{E}_\tau^2(\lambda_{2,\tau}, \mathsf{r}_2, c_2)$,

$$\left|\ell_T(\tau, \hat{\theta}_{1,\tau}, \hat{\theta}_{2,\tau}) - \ell_T(\tau, \theta_\star^{(1)}, \theta_\star^{(2)})\right| \le \frac{\delta(\tau)}{T},$$

where

$$\delta(\tau) \stackrel{\text{def}}{=} \Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)\left[2s_1^{1/2}T\lambda_{1,\tau} + \frac{\tau c_1}{2}\Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)\right]$$
$$+\Psi_{\mathsf{r}_2}\left(6\left(\frac{T}{T-\tau}\right)s_2^{1/2}\lambda_{2,\tau}\right)\left[2s_2^{1/2}T\lambda_{2,\tau} + \frac{(T-\tau)c_2}{2}\Psi_{\mathsf{r}_2}\left(6\left(\frac{T}{T-\tau}\right)s_2^{1/2}\lambda_{2,\tau}\right)\right].$$

PROOF.

$$\ell_T(\tau, \hat{\theta}_{1,\tau}, \hat{\theta}_{2,\tau}) - \ell_T(\tau, \theta_\star^{(1)}, \theta_\star^{(2)}) = \frac{1}{T}\sum_{t=1}^{\tau}\left[\phi_t(\hat{\theta}_{1,\tau}, X^{(t)}) - \phi_t(\theta_\star^{(1)}, X^{(t)})\right]$$
$$+ \frac{1}{T}\sum_{t=\tau+1}^{T}\left[\phi_t(\hat{\theta}_{2,\tau}, X^{(t)}) - \phi_t(\theta_\star^{(2)}, X^{(t)})\right].$$

From the definition

$$\frac{1}{T}\sum_{t=1}^{\tau}\left[\phi_t(\hat{\theta}_{1,\tau}, X^{(t)}) - \phi_t(\theta_\star^{(1)}, X^{(t)})\right] = \left\langle G_\tau^1, \hat{\theta}_{1,\tau} - \theta_\star^{(1)}\right\rangle + \mathcal{L}_1(\tau, \hat{\theta}_{1,\tau}).$$

On $\mathcal{E}_\tau^1(\lambda_{1,\tau}, \mathsf{r}_1, c_1)$, and using Lemma 1, we have

$$\left|\left\langle G_\tau^1, \hat{\theta}_{1,\tau} - \theta_\star^{(1)}\right\rangle\right| \le \frac{\lambda_{1,\tau}}{2}\|\hat{\theta}_{1,\tau} - \theta_\star^{(1)}\|_1 \le 2s_1^{1/2}\lambda_{1,\tau}\Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right),$$

and

$$\mathcal{L}_1(\tau, \hat{\theta}_{1,\tau}) \le \frac{\tau}{T}\frac{c_1}{2}\|\hat{\theta}_{1,\tau} - \theta_\star^{(1)}\|_2^2 \le \frac{\tau c_1}{2T}\Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)^2.$$

Hence

$$\left|\frac{1}{T}\sum_{t=1}^{\tau}\left[\phi_t(\hat{\theta}_{1,\tau}, X^{(t)}) - \phi_t(\theta_\star^{(1)}, X^{(t)})\right]\right|$$
$$\le \frac{1}{T}\Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)\left[2s_1^{1/2}T\lambda_{1,\tau} + \frac{\tau c_1}{2}\Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)\right].$$

A similar bound holds for the second term, and the lemma follows easily. □

We are now in position to prove Theorem S1. We have

$$\mathbb{P}\left(|\hat{\tau} - \tau_\star| > B\right) = \mathbb{P}\left(\hat{\tau} > \tau_\star + B\right) + \mathbb{P}\left(\hat{\tau} < \tau_\star - B\right).$$

We bound the first term $\mathbb{P}(\hat{\tau} > \tau_\star + B)$. The second term follows similarly by working with the reversed sequence $X^{(T)}, \ldots, X^{(1)}$.

For $\tau > \tau_\star$, we shall use $\ell_T(\tau)$ instead of $\ell_T\left(\tau; \hat{\theta}_{1,\tau}, \hat{\theta}_{2,\tau}\right)$ for notational convenience, and we define $r_T(\tau) \stackrel{\text{def}}{=} \ell_T(\tau) - \ell_T\left(\tau, \theta_\star^{(1)}, \theta_\star^{(2)}\right)$. We have

$$\begin{aligned}
\ell_T(\tau) &= \ell_T\left(\tau, \theta_\star^{(1)}, \theta_\star^{(2)}\right) + r_T(\tau), \\
&= \left[\ell_T\left(\tau, \theta_\star^{(1)}, \theta_\star^{(2)}\right) - \ell_T\left(\tau_\star, \theta_\star^{(1)}, \theta_\star^{(2)}\right)\right] + \ell_T\left(\tau_\star, \theta_\star^{(1)}, \theta_\star^{(2)}\right) + r_T(\tau).
\end{aligned}$$

Hence

$$\ell_T(\tau) - \ell_T(\tau_\star) = \left[\ell_T\left(\tau, \theta_\star^{(1)}, \theta_\star^{(2)}\right) - \ell_T\left(\tau_\star, \theta_\star^{(1)}, \theta_\star^{(2)}\right)\right] + r_T(\tau) - r_T(\tau_\star). \tag{S7}$$

It is straightforward to check that for $\tau > \tau_\star$,

$$\ell_T\left(\tau, \theta_\star^{(1)}, \theta_\star^{(2)}\right) - \ell_T\left(\tau_\star, \theta_\star^{(1)}, \theta_\star^{(2)}\right) = \frac{1}{T}\sum_{t=\tau_\star+1}^{\tau}\left(\phi_t(\theta_\star^{(1)}, X^{(t)}) - \phi_t(\theta_\star^{(2)}, X^{(t)})\right).$$

Therefore, and using the definition of $U^{(t)}$ and $\kappa_0^{(t)}$, (S7) becomes

$$\ell_T(\tau) - \ell_T(\tau_\star) = \frac{1}{T}\sum_{t=\tau_\star+1}^{\tau}\kappa_0^{(t)} + \frac{1}{T}\sum_{t=\tau_\star+1}^{\tau}U^{(t)} + r_T(\tau) - r_T(\tau_\star). \tag{S8}$$

We conclude from Lemma 2 that on the event $\mathcal{E}$,

$$\ell_T(\tau) - \ell_T(\tau_\star) = \frac{1}{T}\sum_{t=\tau_\star+1}^{\tau}\kappa_0^{(t)} + \frac{1}{T}\sum_{t=\tau_\star+1}^{\tau}U^{(t)} + \epsilon_T(\tau),$$
$$\text{where } |\epsilon_T(\tau)| \leq \frac{2\sup_{\tau\mathcal{T}}|\delta(\tau)|}{T} = \frac{2\delta}{T}. \tag{S9}$$

Therefore,

$$\mathbb{P}(\hat{\tau} > \tau + B) \leq \mathbb{P}(\mathcal{E}^c) + \sum_{j\geq 0,\ \tau_\star+\lceil B\rceil+j\in\mathcal{T}}\mathbb{P}\left(\mathcal{E},\ \hat{\tau} = \tau_\star + \lceil B\rceil + j\right).$$

Using (S9), we have

$$\begin{aligned}
\mathbb{P}\left(\mathcal{E},\ \hat{\tau} = \tau_\star + \lceil B\rceil + j\right) &\leq \mathbb{P}\left(\mathcal{E},\ \ell_T(\tau_\star + \lceil B\rceil + j) \leq \ell_T(\tau_\star)\right) \\
&\leq \mathbb{P}\left(\left|\sum_{t=\tau_\star+1}^{\tau_\star+\lceil B\rceil+j}U^{(t)}\right| > \sum_{t=\tau_\star+1}^{\tau_\star+\lceil B\rceil+j}\kappa_0^{(t)} - 2\delta\right).
\end{aligned}$$

However, since $B > B_0$, by Assumption A1,

$$\sum_{t=\tau_\star+1}^{\tau_\star+\lceil B\rceil+j} \kappa_0^{(t)} - 2\delta \geq (\lceil B\rceil + j)\,\bar{\kappa}_0\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2 - 2\delta \geq \frac{1}{2}\,(\lceil B\rceil + j)\,\bar{\kappa}_0\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2.$$

The first part of A1 implies that the random variables $Z^{(t)}$ are sub-Gaussian, and by standard exponential bounds for sub-Gaussian random variables, we then have

$$\begin{aligned} \mathbb{P}\left[\mathcal{E},\ \ell_T(\tau_\star + \lceil B\rceil + j) \leq \ell_T(\tau_\star)\right] \quad &\leq \quad 2\exp\left(-\frac{(\lceil B\rceil + j)^2\,\bar{\kappa}_0^2\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^4}{8\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2\sum_{t=\tau_\star+1}^{\tau_\star+\lceil B\rceil+j}\sigma_{0t}^2}\right), \\ &\leq \quad 2\exp\left(-\frac{(\lceil B\rceil + j)\,\bar{\kappa}_0^2\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2}{8\bar{\sigma}_0^2}\right), \end{aligned}$$

where the last inequality uses (S4). We can conclude that

$$\begin{aligned} \mathbb{P}\left[\hat{\tau} > \tau_\star + B\right] \quad &\leq \quad \mathbb{P}(\mathcal{E}^c) + 2\sum_{j\geq 0}\exp\left(-\frac{(\lceil B\rceil + j)\,\bar{\kappa}_0^2\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2}{8\bar{\sigma}_0^2}\right) \\ &\leq \quad \mathbb{P}(\mathcal{E}^c) + 2\frac{\exp\left(-\frac{B\bar{\kappa}_0^2\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2}{8\bar{\sigma}_0^2}\right)}{1 - \exp\left(-\frac{\bar{\kappa}_0^2\|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2}{8\bar{\sigma}_0^2}\right)}, \end{aligned} \tag{S10}$$

as claimed. □

S2. Proof of Theorem 1

We will deduce Theorem 1 from Theorem S1. We take Θ as $\mathcal{M}_p$, the set of all $p \times p$ real symmetric matrices, equipped with the (modified) Frobenius inner product $\langle\theta, \vartheta\rangle_{\mathsf{F}} \stackrel{\text{def}}{=} \sum_{k\leq j}\theta_{jk}\vartheta_{jk}$, and the associated norm $\|\theta\|_{\mathsf{F}} \stackrel{\text{def}}{=} \sqrt{\langle\theta, \theta\rangle}$. With this inner product, we identify $\mathcal{M}_p$ with the Euclidean space $\mathbb{R}^d$, with $d = p(p+1)/2$. This puts us in the setting of Theorem S1.

We will use the following notation. If $u \in \mathbb{R}^q$, for some integer $q \geq 1$, and A is an ordered subset of $\{1, \ldots, q\}$, we define $u_A \stackrel{\text{def}}{=} (u_j,\ j \in A)$, and u_{-j} is a shortcut for $u_{\{1,\ldots,q\}\setminus\{j\}}$. We define the function $B_{jk}(x, y) = B_0(x)$ if $j = k$, and $B_{jk}(x, y) = B(x, y)$ if $j \neq k$.

In the present case, the function ϕ_t is ϕ as given in (5), and does not depend on t. The following properties of the conditional distribution (3) will be used below. It is well known (and easy to prove using Fisher's identity) that the function $\theta \mapsto \phi(\theta, x)$

is Lispchitz and

$$|\phi(\theta, x) - \phi(\vartheta, x)| \leq 2c_0 \|\theta - \vartheta\|_1, \;\; \theta, \vartheta \in \mathcal{M}_p, \; x \in \mathsf{X}^p, \tag{S11}$$

where c_0 is as in (9). From the expression (3) of the conditional densities, using straightforward algebra, it is easy to show that the negative log-pseudo-likelihood function $\phi(\theta, x)$ satisfies the following. For all $\theta, \Delta \in \mathcal{M}_p$, and $x \in \mathsf{X}^p$,

$$\begin{aligned} &\phi(\theta + \Delta, x) - \phi(\theta, x) - \langle \nabla_\theta \phi(\theta, x), \Delta \rangle_{\mathsf{F}} \\ &\quad = \sum_{j=1}^{p} \left[\log Z_{\theta+\Delta}^{(j)}(x) - \log Z_{\theta}^{(j)}(x) - \sum_{k=1}^{p} \Delta_{jk} \frac{\partial}{\partial \theta_{jk}} \log Z_{\theta}^{(j)}(x) \right]. \end{aligned} \tag{S12}$$

Furthermore by Taylor expansion, we have

$$\begin{aligned} &\log Z_{\theta+\Delta}^{(j)}(x) - \log Z_{\theta}^{(j)}(x) - \sum_{k=1}^{p} \Delta_{jk} \frac{\partial}{\partial \theta_{jk}} \log Z_{\theta}^{(j)}(x) \\ &= \int_0^1 (1-t) \mathsf{Var}_{\theta + t\Delta} \left(\sum_{k=1}^{p} \Delta_{jk} B_{jk}(X_j, X_k) | X_{-j} \right) \mathrm{d}t \leq \frac{c_0^2}{2} \left(\sum_{k=1}^{p} |\Delta_{jk}| \right)^2. \end{aligned} \tag{S13}$$

By the self-concordant bound derived in Atchadé (2014) Lemma A2, we have

$$\begin{aligned} &\log Z_{\theta+\Delta}^{(j)}(x) - \log Z_{\theta}^{(j)}(x) - \sum_{k=1}^{p} \Delta_{jk} \frac{\partial}{\partial \theta_{jk}} \log Z_{\theta}^{(j)}(x) \\ &\qquad \geq \frac{1}{2 + c_0 \sum_{k=1}^{p} |\Delta_{jk}|} \mathsf{Var}_{\theta} \left(\sum_{k=1}^{p} \Delta_{jk} B_{jk}(X_j, X_k) | X_{-j} \right). \end{aligned} \tag{S14}$$

Proof (Proof of Theorem 1). Let us first show that under assumption H3 of Theorem 1, A1 holds. Since in this case ϕ_t does not actually depend on t, we can take $B_0 = 1$ in A1, and (S3) follows automatically from H3 with $\bar{\kappa}_0 = \kappa / \|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2$. Also, (S11) implies that $|U^{(t)}| \leq 4c_0 \|\theta_\star^{(2)} - \theta_\star^{(1)}\|_1 \leq 4c_0 s^{1/2} \|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2$, where s denotes the number of non-zero entries of $\theta^{(2)} - \theta_\star^{(1)}$. Hence for all $x > 0$,

$$\mathbb{E}\left(e^{xU^{(t)}} \right) \leq \exp\left(8x^2 c_0^2 s \|\theta_\star^{(2)} - \theta_\star^{(1)}\|_2^2 \right).$$

This establishes the sub-Gaussian condition of A1, and (S4) holds with $\bar{\sigma}_0^2 = 16 c_0^2 s$.

For $j = 1, 2$, let $\lambda_{1,\tau}$, $\lambda_{2,\tau}$ as in (8). We will apply Theorem S1 with $c_j = 64 c_0 s_j$, the rate function $\mathsf{r}_j(x) = \frac{\rho_j x^2}{2 + 4c_0 s_j^{1/2} x}$, $x > 0$, and with the event $\mathcal{E} =$

$\bigcap_{\tau\in\mathcal{T}}\left[\mathcal{E}_\tau^1\left(\lambda_{1,\tau},\mathsf{r}_1,c_1\right)\cap\mathcal{E}_\tau^2\left(\lambda_{2,\tau},\mathsf{r}_2,c_2\right)\right]$, where the search domain $\mathcal{T}$ satisfies (15), (16), and (18). Notice that if $\mathsf{r}(x)=\rho x^2/(2+bx)$, $\rho,b>0$, is a rate function, then for $a>0$, $\Psi_{\mathsf{r}}(a)\stackrel{\text{def}}{=}\inf\{x>0:\ r(x)\geq ax\}\leq 4a/\rho$, provided that $2ba\leq\rho$. Hence

$$\Psi_{\mathsf{r}_1}\left(6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}\right)\leq\frac{4}{\rho_1}6\left(\frac{T}{\tau}\right)s_1^{1/2}\lambda_{1,\tau}=24\times 32c_2\frac{s_1^{1/2}}{\rho_1}\sqrt{\frac{\log(dT)}{\tau}},$$

provided that $\tau\geq(48\times 32)^2c_0^2\left(\frac{s_1}{\rho_1}\right)^2\log(dT)$. Therefore, given that all $\tau\in\mathcal{T}$ satisfies (18), with some simple algebra we see that there exists a universal constant a that we can take as $a=(24\times 32\times 64)^2$, such that for all $\tau\in\mathcal{T}$,

$$\delta(\tau)\leq\delta=ac_0^2M\log(dT),$$

where

$$M=\left[\frac{s_1}{\rho_1}\left(1+c_0\frac{s_1}{\rho_1}\right)+\frac{s_2}{\rho_2}\left(1+c_0\frac{s_2}{\rho_2}\right)\right].$$

Therefore in Theorem S1, we can take $B=\frac{4ac_0^2M\log(dT)}{\kappa}$, and by the conclusion of Theorem S1,

$$\mathbb{P}\left[|\hat{\tau}-\tau_\star|>B\right]\leq 2\mathbb{P}(\mathcal{E}^c)+\frac{4\exp\left(-\frac{\delta}{32c_0^2s}\left(\frac{\kappa}{\|\theta_\star^{(2)}-\theta_\star^{(1)}\|_2^2}\right)^2\right)}{1-\exp\left(-\frac{\kappa^2}{2^7c_0^2s\|\theta_\star^{(2)}-\theta_\star^{(1)}\|_2^2}\right)}.$$

We show in Lemma 3 and Lemma 4 below that $\mathbb{P}(\mathcal{E}^c)\leq 8/d$, and this ends the proof.

$\square$

LEMMA 3. *Let $\lambda_{1,\tau},\lambda_{2,\tau}$ be as in equation* (8). *Suppose that the search domain $\mathcal{T}$ is such that (15)-(16) hold. Then*

$$\mathbb{P}\left[\max_{\tau\in\mathcal{T}}\lambda_{1,\tau}^{-1}\left\|G_\tau^1\right\|_\infty>\frac{1}{2}\right]\leq\frac{2}{d},\quad\textit{and}\quad\mathbb{P}\left[\max_{\tau\in\mathcal{T}}\lambda_{2,\tau}^{-1}\left\|G_\tau^2\right\|_\infty>\frac{1}{2}\right]\leq\frac{2}{d},$$

where $d=p(p+1)/2$.

PROOF. We carry the details for the first bound. The second is done similarly by working with the reversed sequence $X^{(T)},\ldots,X^{(1)}$. Fix $1\leq j\leq i\leq p$, $t\in\mathcal{T}$, and define $V_{ij}^{(t)}\stackrel{\text{def}}{=}\frac{\partial}{\partial\theta_{ij}}\phi(\theta_\star^{(1)},X^{(t)})$. We calculate that

$$V_{ij}^{(t)}=\begin{cases}-B_0(X_i^{(t)})+\mathbb{E}_{\theta_\star^{(1)}}(B_0(X_i|X_{-i}^{(t)}) & \text{if } i=j\\ -2B(X_i^{(t)},X_j^{(t)})+\mathbb{E}_{\theta_\star^{(1)}}\left(B(X_i,X_j^{(t)})|X_{-i}^{(t)}\right)+\mathbb{E}_{\theta_\star^{(1)}}\left(B(X_i,X_j^{(t)})|X_{-j}^{(t)}\right) & \text{if } j<i.\end{cases}$$

In the above display the notation $\mathbb{E}_{\theta_\star^{(1)}}\left(B(X_i, X_j^{(t)})|X_{-i}^{(t)}\right)$ is defined as the function $z \mapsto \mathbb{E}_{\theta_\star^{(1)}}\left(B(X_i, z_j)|X_{-i} = z_{-i}\right)$ evaluated on $X^{(t)}$. Since $X^{(1:\tau_\star)} \overset{i.i.d}{\sim} g_{\theta_\star^{(1)}}$, it follows that $\mathbb{E}(V_{ij}^{(t)}) = 0$ for $t = 1, \ldots, \tau_\star$. We set $\mu_{ij} \overset{\text{def}}{=} \mathbb{E}(V_{ij}^{(\tau_\star+1)}) = \mathbb{E}(V_{ij}^{(t)})$ for $t = \tau_\star + 1, \ldots, T$. We also set $\bar{V}_{ij}^{(t)} \overset{\text{def}}{=} V_{ij}^{(t)} - \mathbb{E}\left(V_{ij}^{(t)}\right)$. It is easy to see that $|\bar{V}_{ij}^{(t)}| \leq 4c_0$, where c_0 is defined in (9) . With these notations, for $\tau \in \mathcal{T}$, we can write

$$(G_\tau^1)_{ij} = \frac{1}{T}\sum_{t=1}^{\tau} \bar{V}_{ij}^{(t)} + \frac{(\tau - \tau_\star)_+ \mu_{ij}}{T},$$

where $a_+ \overset{\text{def}}{=} \max(a, 0)$. For $t > \tau_\star$, Lemma 5 can be used to write

$$\begin{aligned}
&\left|\mathbb{E}\left[B(X_i^{(t)}, X_j^{(t)}) - \mathbb{E}_{\theta_\star^{(1)}}\left(B(X_i, X_j^{(t)})|X_{-i}^{(t)}\right)\right]\right| \\
&\quad = \left|\mathbb{E}\left[\int_{\mathsf{X}} B(u, X_j^{(t)}) f_{\theta_\star^{(2)}}(u|X_{-i}^{(t)})\mathrm{d}u - \int_{\mathsf{X}} B(u, X_j^{(t)}) f_{\theta_\star^{(1)}}(u|X_{-i}^{(t)})\mathrm{d}u\right]\right| \\
&\quad\quad \leq c_0^2 \sum_{j=1}^{p} |\theta_{\star,ij}^{(2)} - \theta_{\star,ij}^{(1)}| \leq bc_0^2,
\end{aligned}$$

where b is as in (17). Hence

$$|\mu_{ij}| \leq 2\max_{j \leq i}\left|\mathbb{E}_{\theta_\star^{(2)}}\left[B(X_i^{(t)}, X_j^{(t)}) - \mathbb{E}_{\theta_\star^{(1)}}\left(B(X_i^{(t)}, X_j^{(t)})|X_{-j}^{(t)}\right)\right]\right| \leq 2bc_0^2.$$

Set $\lambda_\tau \overset{\text{def}}{=} (A\sqrt{\tau}/T)$, where

$$A \overset{\text{def}}{=} 32c_0\sqrt{\log(dT)}.$$

By a union-bound argument,

$$\begin{aligned}
&\mathbb{P}\left[\max_{\tau \in \mathcal{T}} 2\lambda_\tau^{-1}\|G_\tau^1\|_\infty > 1\right] \\
&\quad \leq \sum_{\tau \in \mathcal{T}}\sum_{i,j} \mathbb{P}\left[\frac{1}{A\sqrt{\tau}}\left|\sum_{t=1}^{\tau} \bar{V}_{ij}^{(t)}\right| + \frac{2bc_0^2(\tau - \tau_\star)_+}{A\sqrt{\tau}} > \frac{1}{2}\right]. \quad \text{(S15)}
\end{aligned}$$

Since $A = 32c_0\sqrt{\log(dT)}$, for $\tau \in \mathcal{T}$, and using (15) we see that $\max_{\tau \in \mathcal{T}} \frac{2bc_0^2(\tau - \tau_\star)_+}{A\sqrt{\tau}} \leq 1/4$. Hence

$$\begin{aligned}
\mathbb{P}\left[\max_{\tau \in \mathcal{T}} 2\lambda_\tau^{-1}\|G_\tau^1\|_\infty > 1\right] &\leq \sum_{\tau \in \mathcal{T}}\sum_{i,j} \mathbb{P}\left[\left|\sum_{t=1}^{\tau} \bar{V}_{ij}^{(t)}\right| > \frac{A\sqrt{\tau}}{4}\right], \quad \text{(S16)} \\
&\leq 2\sum_{\tau \in \mathcal{T}}\sum_{i,j} \exp\left(-\frac{A^2}{8^3 c_0^2}\right) \leq \frac{2}{d}.
\end{aligned}$$

where the second inequality uses Hoeffding's inequality. □

REMARK 1. *The* $\log(dT)$ *term that appears in the convergence rate of Theorem 1 follows from the union bound and the exponential bound used in (S15), and (S16) respectively. Alternatively, it is easy to see that one could also write*

$$\mathbb{P}\left[\max_{\tau\in\mathcal{T}} 2\lambda_\tau^{-1}\|G_\tau^1\|_\infty > 1\right] \leq \sum_{i,j}\mathbb{P}\left[\max_{\tau\in\mathcal{T}}\left|\frac{1}{\sqrt{\tau}}\sum_{t=1}^{\tau}\bar{V}_{ij}^{(t)}\right| > \frac{A}{4}\right].$$

Hence whether one can remote the $\log(T)$ *term hinges on the existence of an exponential bound for the term* $\max_{\tau\in\mathcal{T}}\left|\tau^{-1/2}\sum_{t=1}^{\tau}\bar{V}_{ij}^{(t)}\right|$. *Unfortunately we are not aware of any such result in the literature. The closest results available deal with the unweighted sums:* $\max_{\tau\in\mathcal{T}}\left|\sum_{t=1}^{\tau}\bar{V}_{ij}^{(t)}\right|$ *(see for instance pinelis (2006) for some of the best bounds available).*

LEMMA 4. *Assume H1 and H2. Let* $\lambda_{1,\tau}$ *and* $\lambda_{2,\tau}$ *as in Equation* (8), *and let the search domain* $\mathcal{T}$ *be such that Equations* (15)-(16) *hold. Take* $c_1 = 64c_0s_1$, $c_2 = 64c_0s_2$ *and*

$$\mathsf{r}_1(x) = \frac{\rho_1 x^2}{2+4c_0 s_1^{1/2}x}, \quad \text{and} \quad \mathsf{r}_2(x) = \frac{\rho_2 x^2}{2+4c_0 s_2^{1/2}x}, \quad x \geq 0.$$

Then the event $\bigcap_{\tau\in\mathcal{T}}\left[\mathcal{E}_\tau^1\left(\lambda_{1,\tau},\mathsf{r}_1,c_1\right)\cap\mathcal{E}_\tau^2\left(\lambda_{2,\tau},\mathsf{r}_2,c_2\right)\right]$ *holds with probability at least* $1-\frac{8}{d}$.

PROOF. We have seen in Lemma 3 that with $\lambda_{1,\tau}$ and $\lambda_{2,\tau}$ as in equation (8), the event $\cap_{\tau\in\mathcal{T}}\left[\{\|G_\tau^1\|_\infty \leq \lambda_{1,\tau}/2\}\cap\{\|G_\tau^1\|_\infty \leq \lambda_{2,\tau}/2\}\right]$ holds with probability at least $1-2/d$. We have

$$\mathcal{L}_1(\tau,\theta) \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{\tau}\left[\phi(\theta,X^{(t)}) - \phi(\theta_\star^{(1)},X^{(t)}) - \left\langle\nabla\phi(\theta_\star^{(1)},X^{(t)}),\theta-\theta_\star^{(1)}\right\rangle\right].$$

(S13) then implies that for all $\tau\in\mathcal{T}$, and $\theta-\theta_\star^{(1)}\in\mathbb{C}_1$,

$$\mathcal{L}_1(\tau,\theta) \leq \frac{\tau}{T}\frac{4c_0^2}{2}\|\theta-\theta_\star^{(1)}\|_1^2 \leq \frac{\tau}{T}\frac{64c_0^2 s_1}{2}\|\theta-\theta_\star^{(1)}\|_2^2.$$

A similar bound holds for $j=2$. Hence $\cap_{\tau\in\mathcal{T}}\cap_{j=1}^2\left\{\sup_{\theta\neq\theta_\star^{(j)},\,\theta-\theta_\star^{(j)}\in\mathbb{C}_j}\frac{\mathcal{L}_j(\tau,\theta)}{\|\theta-\theta_\star^{(j)}\|_2^2} \leq \frac{\tau}{T}\frac{c_j}{2}\right\}$ holds with probability one.

Using (S14), we have

$$\mathcal{L}_1(\tau,\theta) \geq \frac{\tau}{T}\frac{1}{2+4c_0 s_1^{1/2}\|\theta-\theta_\star^{(1)}\|_2} \times\frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^{p}\mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^{p}B_{kj}(X_j^{(t)},X_k^{(t)})\left(\theta_{kj}-\theta_{\star,kj}^{(1)}\right)|X_{-j}^{(t)}\right). \quad \text{(S17)}$$

We will now show that for all $\tau \in \mathcal{T}$, and all $\theta - \theta_\star^{(1)} \in \mathbb{C}_1$, with probability at least $1 - 2/d$, we have

$$\frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^{p} \mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^{p} B_{kj}(X_j^{(t)}, X_k^{(t)})\left(\theta_{kj} - \theta_{\star,kj}^{(1)}\right)|X_{-j}^{(t)}\right) \geq \rho_1\|\theta - \theta_\star^{(1)}\|_2^2.$$

Given (S17), this assertion will implies that $\mathcal{L}_1(\tau, \theta) \geq \frac{\tau}{T}\mathsf{r}_1(\|\theta - \theta_\star^{(1)}\|_2)$ for all $\theta - \theta_\star^{(1)} \in \mathbb{C}_1$ with probability at least $1 - 2/d$, where $\mathsf{r}_1(x) = \rho_1 x^2/(2 + 4c_0 s_1^{1/2} x)$. The lemma will then follow easily.

For $\Delta \in \mathcal{M}_p$, we define

$$\mathcal{V}^1(\tau, \Delta) \stackrel{\text{def}}{=} \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^{p} \mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^{p} B_{kj}(X_j^{(t)}, X_k^{(t)})\Delta_{kj}|X_{-j}^{(t)}\right),$$

and

$$W_{jkk'}^{(t)} \stackrel{\text{def}}{=} \mathsf{Cov}_{\theta_\star^{(1)}}\left(B(X_j^{(t)}, X_k^{(t)}), B(X_j^{(t)}, X_{k'}^{(t)})|X_{-j}^{(t)}\right) - \mathbb{E}\left[\mathsf{Cov}_{\theta_\star^{(1)}}\left(B(X_j^{(t)}, X_k^{(t)}), B(X_j^{(t)}, X_{k'}^{(t)})|X_{-j}^{(t)}\right)\right].$$

Then for $\Delta \in \mathbb{C}_1 \setminus \{0\}$,

$$\begin{aligned}\mathcal{V}^1(\tau, \Delta) &= \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^{p}\sum_{k,k'=1}^{p} \Delta_{jk}\Delta_{jk'}\mathbb{E}\left[\mathsf{Cov}_{\theta_\star^{(1)}}\left(B(X_j^{(t)}, X_k^{(t)}), B(X_j^{(t)}, X_{k'}^{(t)})|X_{-j}^{(t)}\right)\right]. \\ &\quad + \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^{p}\sum_{k,k'=1}^{p} \Delta_{jk}\Delta_{jk'}W_{jkk'}^{(t)} \qquad \text{(S18)}\end{aligned}$$

Using H1, we deduce that

$$\begin{aligned}\mathcal{V}^1(\tau, \Delta) \geq 2\rho_1\|\Delta\|_2^2 &+ \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^{p}\sum_{k,k'=1}^{p} \Delta_{jk}\Delta_{jk'}W_{jkk'}^{(t)} \\ &+ \frac{(\tau - \tau_\star)_+}{\tau}\sum_{j=1}^{p}\mathbb{E}_{\theta_\star^{(2)}}\left[\mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^{p}\Delta_{jk}B_{ik}(X_j, X_k)|X_{-j}\right)\right] \\ &- \frac{(\tau - \tau_\star)_+}{\tau}\sum_{j=1}^{p}\mathbb{E}_{\theta_\star^{(1)}}\left[\mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^{p}\Delta_{jk}B_{ik}(X_j, X_k)|X_{-j}\right)\right]. \qquad \text{(S19)}\end{aligned}$$

By the comparison Lemma 5

$$\left|\mathbb{E}_{\theta_\star^{(2)}}\left[\mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^p \Delta_{jk}B_{ik}(X_j,X_k)|X_{-j}\right)\right] - \mathbb{E}_{\theta_\star^{(1)}}\left[\mathsf{Var}_{\theta_\star^{(1)}}\left(\sum_{k=1}^p \Delta_{jk}B_{ik}(X_j,X_k)|X_{-j}\right)\right]\right|$$
$$\leq c_0^3\left(\sum_{k=1}^p|\Delta_{jk}|\right)^2\sum_{k=1}^p|\theta_{\star jk}^{(1)}-\theta_{\star jk}^{(2)}| \leq c_0^3 b\left(\sum_{k=1}^p|\Delta_{jk}|\right)^2,$$

which implies that

$$\mathcal{V}^1(\tau,\Delta) \geq \left(2\rho_1 - \frac{64}{\tau}(\tau-\tau_\star)_+ s_1 c_0^3 b\right)\|\Delta\|_2^2 + \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^p\sum_{k,k'=1}^p \Delta_{jk}\Delta_{jk'}W_{jkk'}^{(t)}.$$

Given that on $\mathcal{T}_+$, $128(\tau-\tau_\star)s_1c_0^3b \leq \rho_1\tau$, it follows that for all $\tau\in\mathcal{T}$,

$$\mathcal{V}^1(\tau,\Delta) \geq \frac{3}{2}\rho_1\|\Delta\|_2^2 + \frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{j=1}^p\sum_{k,k'=1}^p \Delta_{jk}\Delta_{jk'}W_{jkk'}^{(t)} \tag{S20}$$

Set $Z_{jkk'}^{\tau} \stackrel{\text{def}}{=} \frac{1}{\tau}\sum_{t=1}^{\tau}W_{jkk'}^{(t)}$. We conclude from equation (S20) that if for some $\Delta\in\mathbb{C}_1\setminus\{0\}$, and for some $\tau\in\mathcal{T}$,

$$\mathcal{V}^1(\tau,\Delta) \leq \rho_1\|\Delta\|_2^2 \tag{S21}$$

then

$$\sum_{j=1}^p\sum_{k,k'=1}^p \Delta_{jk}\Delta_{jk'}Z_{jkk'}^{(\tau)} \leq -\frac{\rho_1}{2}\|\Delta\|_2^2.$$

But on the other hand, using the fact that $\Delta\in\mathbb{C}_1$,

$$\begin{aligned}\sum_{j=1}^p\sum_{k,k'=1}^p \Delta_{jk}\Delta_{jk'}Z_{jkk'}^{(\tau)} &\geq -\left(\sup_{j,k,k'}|Z_{jkk'}^{(\tau)}|\right)\left(\sum_{i=1}^p\sum_{k=1}^p|\Delta_{ik}|\right)^2\\ &\geq -\left(\sup_{j,k,k'}|Z_{jkk'}^{(\tau)}|\right)4\|\Delta\|_1^2\\ &\geq -64s_1\left(\sup_{j,k,k'}|Z_{jkk'}^{(\tau)}|\right)\|\Delta\|_2^2.\end{aligned}$$

Therefore if there exists a non-zero $\Delta\in\mathbb{C}_1$ and $\tau\in\mathcal{T}$ such that equation (S21) holds then $\left(\sup_{j,k,k'}|Z_{jkk'}^{(\tau)}|\right) \geq (\rho_1/s_1)(1/128)$. But by Hoeffding's inequality and a union-sum bound,

$$\mathbb{P}\left[\sup_{j,k,k'} |Z^{(\tau)}_{jkk'}| \geq \frac{\rho_1}{128 s_1}\right] \leq 2\exp\left(3\log p - \frac{\tau\rho_1^2}{2^9 c_0^2 s_1^2}\right) \leq \frac{2}{p},$$

since for $\tau \in \mathcal{T}$, $\tau \geq 2^{11} c_0^2 s_1^2 \rho_1^{-2} \log p$. □

LEMMA 5. *Let* $(\mathsf{Y}, \mathcal{A}, \nu)$ *be a measure space where* ν *is a finite measure. Let* $g_1, g_2, f_1, f_2 : \mathsf{Y} \to \mathbb{R}$ *be bounded measurable functions. Set* $Z_{g_i} \stackrel{\text{def}}{=} \int_{\mathsf{Y}} e^{g_i(y)}\nu(dy)$, $i \in \{1, 2\}$. *Then*

$$\left|\frac{1}{Z_{g_1}} \int f_1(y) e^{g_1(y)}\nu(dy) - \frac{1}{Z_{g_2}} \int f_2(y) e^{g_2(y)}\nu(dy)\right| \\ \leq \|f_2 - f_1\|_\infty + \frac{1}{2}\mathsf{osc}(g_2 - g_1)\left(\mathsf{osc}(f_1) + \mathsf{osc}(f_2)\right),$$

where $\|f\|_\infty = \sup_{x\in\mathsf{Y}} |f(x)|$, *and* $\mathsf{osc}(f) \stackrel{\text{def}}{=} \sup_{x,y\in\mathsf{Y}} |f(x) - f(y)|$ *is the oscillation of* f.

PROOF. The proof follows from Atchadé (2014) Lemma 3.4.

S3. Different Methods of Missing Data Imputation for the Real Data Application

In the main paper we replaced the missing votes by the value (yes/no) of that member's party majority position on that particular vote. Here we employed two other missing data imputation techniques viz. (i) replacing all missing values by the value (yes/no) representing the winning majority on that bill and (ii) replacing the missing value of a Senator by the value that the majority of the opposite party voted on that particular bill. The estimated change-point obtained following these two imputation methods are not much different . The imputation technique (i) results in a estimated change-point at January 19, 1995 and the technique (ii) yields estimated change-point at January 17, 1995 respectively. The change-point estimate we obtained in the main paper was January 17, 1995. Clearly there is not much difference between the different imputation techniques and Fig. S1 also conveys the same message.

Fig. S1: Estimated Change-points via imputation technique (i) and (ii) respectively

References

Atchadé, F. Y. (2014). Estimation of Network Structures from partially observed markov random field. *Electron. J. Statist.*, **8**, 2242-2263

Bach, F. (2010). Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, **4**, 384-414.

Bai, J.(2010). Estimation of a change-point in multiple regression models. *The Review of Economics and Statistics*, **4**, 551-563.

Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008) Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J.Mach.Learn.Res.*, **9**, 485-516.

Basu, S. and Michailidis, G. (2015). Estimation in high dimensional vector autoregressive models. *Ann. Statist.* To Appear.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B.*, **36**, 192-236.

Bhattcharya, K. P. (1974). Maximum likelihood estimation of a change-point in the distribution of the independent random variables: General Multiparameter Case*J. Mult. Anls.*, **23**, 183-208.

Bickel, P.J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, **36**, 199-227.

Bickel, P. J., Ritov, Y. and Tsybakov, A.B. (2009) Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.*, **37**, 1705-1732.

Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.*, **17**, 453-510.

Carlstein, E. (1988). Nonparametric change-point estimation. *Ann. Statist.*, **16**, 188-197.

Drton, M. and Perlman, M.D. (2004). Model selection for Gaussian concentration graphs. *Biometrika.*, **91**, 591-602.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularized paths for generalized linear models via coordinate descent. *J. Statist. Softwr.*, **33**, 1-22

Guo, J., Levina, E., Michailidis, G. and Zhu, J. (2010). Joint structure estimation for categorical markov networks. *Tech. rep.*, Univ. of Michigan.

Han, F. and Liu, H. (2006). A direct estimation of high dimensional stationary vector autoregressions*arXiv:1307.0293v2 [stat.ML]* .

Hanneke, S. and Xing, P. E. (2006). Discrete temporal models of social networks. *Lecture Notes in Computer Science.*, **4503**, 115-125.

Hinkley, V. D. (1970). Inference about the change-point in a sequence of random variables. *Biometrika.*, **57**, 1-17.

Hinkley, V. D. (1972). Time-ordered classification. *Biometrika.*, **59**, 509-523.

Hoefling, H.(2010). BMN: The pseudo-likelihood method for pairwise binary markov networks. *R package version 1.02, http://CRAN.R-project.org/package=BMN.*

Höfling, H. and Tibshirani, R. (2009). Estimation of Sparse Binary Pairwise Markov Networks using Pseudo-likelihoods. *J. Mach. Learn. Res.* **10**, 883-906.

Hurvich, M. C., Simonoff, S.J. and Tsai, C. (1998). Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *J. R. Stat. Soc. Ser. B.*, **60**, 271-293.

Kolar, M., Song, L., Ahmed, A. and Xing, P. E. (2010). Estimating Time varying Networks. *Ann. App. Statist.*, **4**, 94-123.

Kolar, M. and Xing, P. E. (2012). Estimating networks with jumps. *Electron. J. Statist.*, **6**, 2069-2106.

Kosorok, R. M. (2012). Introduction to empirical processes and semiparametric inference. *Springer Series in Statistics.*

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix. *Ann. Statist.*, **37**, 4254-4278.

Lan, Y., Banerjee M. and Michailidis G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.*, **37**, 1752-1791.

Loader C. (1996). Change-point estimation using nonparametric regression. *Ann. Statist.*, **24**, 1667-1678.

Massart, P. (2007). Concentration inequalities and model selection. *Springer Verlag.*

Meinshausen, N. and Bühlmann, P.(2006) High dimensional graphs and variable selection with the lasso. *Ann. Statist.*, **34**, 1436-1462.

Meinshausen, N. and Bühlmann, P.(2010). Stability selection. *J. R. Statist. Soc. B*, **72**, 417-473.

Moody, J. and Mucha, P.(2013). Portrait of political party polarization *Network Science*, **1**, 119-121.

Muller, H.(1992). Change-points in nonparametric regression analysis. *Ann. Statist.*, **20**, 737-761.

Nadaraya, E. A.(1965) On non-parametric estimation of density functions and regression curves. *Theory Prob. Applic.*, **10**, 186-190.

Neghaban, S., Ravikumar, P., Wainwright, M. and Yu, B. (2010). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.*, **27**, 538-557.

Pinelis, I. (2006). On the normal domination of (super) martingales. *Electronic Journal of Probability*, **39**, 1049-1070.

Raimondo, M. (1998). Minimax estimation of sharp change-points. *Ann. Statist.*, **26**, 1379-1397.

Ravikumar, P., Wainwright, J. M. and Lafferty, D. J. (2010). High-dimensional ising model selection using l_1-regularized logistic regression. *Ann. Statist.*, **38**, 1287-1319.

Rothman, A. J., Bickel P. J., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.*, **2**, 494-515.

Schwarz, G.(1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461-464.

Van de Geer, S. A. and Bühlmann, P. (2009). On the conditions used to prove oracle results for the lasso. *Electron. J. Stat.*, **3**, 1360-1392.

Van der Vaart, A. and Wellner, J. (1996). Weak convergence and empirical processes. *Springer Series in Statistics.*

Wainwright, J. M. and Jordan, I. M.(2008) Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning.* **1**, 1-305.

Xue, L., Zou, H. and Cai, T. (2012). Non-concave penalized composite likelihood estimation of sparse ising models. *Ann. Statist.*, **40**, 1403-1429.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika.*, **94**, 19-35.

Zhou, S., Lafferty, J. and Wasserman, L. (2010). Time-varying undirected graphs. *Machine Learning.* , **80**, 295-319.