WEB-BASED SUPPORTING MATERIALS FOR "A FAST ASYNCHRONOUS MCMC SAMPLER FOR SPARSE BAYESIAN INFERENCE", BY YVES ATCHADÉ AND LIWEI WANG

Appendix A. Proof of Theorem 1

Throughout C_0 denotes a generic constant whose value may change from one appearance to the next. We shall write $\Pi(\cdot)$ and $\tilde{\Pi}(\cdot)$ instead of $\Pi(\cdot|\mathcal{D})$ and $\tilde{\Pi}(\cdot|\mathcal{D})$ respectively. For any two Markov kernels P_1 and P_2 and for any integer $k \geq 1$, it is easily checked that

$$P_1^k = P_2^k + \sum_{j=1}^{k} P_1^{k-j}(P_1 - P_2)P_2^{j-1}. \quad (1)$$

Using this identity, for any bounded measurable function $f : \Delta \times \mathbb{R}^p \to \mathbb{R}$, writing $\tilde{f} = f - \tilde{\Pi}(f)$, we have for any $k \geq 0$,

$$\Pi(f) - \tilde{\Pi}(f) = \Pi(\tilde{f}) = \Pi(K^k \tilde{f}) = \Pi(\tilde{K}^k \tilde{f}) + \sum_{j=1}^{k} \Pi\left((K - \tilde{K})\tilde{K}^{j-1}\tilde{f}\right),$$

$$= \Pi(\tilde{K}^k \tilde{f}) + \Pi\left[(K - \tilde{K})\left(\sum_{j=0}^{k-1} \tilde{K}^j \tilde{f}\right)\right]. \quad (2)$$

Define

$$g_k \stackrel{\text{def}}{=} \sum_{j=0}^{k-1} \tilde{K}^j \tilde{f}.$$

It follows from (10) that for all $(\delta, \theta) \in \Delta \times \mathbb{R}^p$,

$$|g_k(\delta, \theta)| \leq C_0 \|f\|_\infty V^{1/2}(\delta, \theta) \sum_{j=0}^{k-1} \tilde{\lambda}^j \leq \frac{C_0 \|f\|_\infty}{1 - \tilde{\lambda}} V^{1/2}(\delta, \theta). \quad (3)$$

Recall that

$$K((\delta, \theta); (\mathrm{d}\delta', \mathrm{d}\theta')) = K_\delta(\theta, \mathrm{d}\theta') \sum_{\mathsf{J}:\ |\mathsf{J}|=J} \binom{p}{J}^{-1} Q_{\theta',\mathsf{J}}(\delta, \mathrm{d}\delta'),$$

where

$$K_\delta(\theta, \mathrm{d}\theta') \stackrel{\text{def}}{=} P_\delta([\theta]_\delta, \mathrm{d}[\theta']_\delta) \prod_{j:\ \delta_j = 0} \mathbf{N}(0, \rho_0^{-1})(\mathrm{d}\theta'_j).$$

$\tilde{K}$ has the same expression, but with Q replaced by $\tilde{Q}$. Therefore, using the fact that K_δ has invariant distribution $\Pi(\cdot|\delta)$, we have

$$\Pi\left((K-\tilde{K})g_k\right) = \sum_{\mathsf{J}:\,|\mathsf{J}|=J} \binom{p}{J}^{-1} \int_{\Delta\times\mathbb{R}^p} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\left\{\int_\Delta Q_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta) - \int_\Delta \tilde{Q}_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta)\right\}$$

Without any loss generality we shall assume now that $\|f\|_\infty = 1$. Using the last display and (2), we get that

$$\left|\Pi(f) - \tilde{\Pi}(f)\right| \leq C_0\tilde{\lambda}^k + \sum_{\mathsf{J}:\,|\mathsf{J}|=J} \binom{p}{J}^{-1} \int_{\Delta\times\mathbb{R}^p} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\left|\int_\Delta Q_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta) - \int_\Delta \tilde{Q}_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta)\right|. \tag{4}$$

We split the integral $\int_{\Delta\times\mathbb{R}^p}$ over B and over B^c. For the part over B^c, we use the Cauchy-Schwarz inequality, and (3) to write

$$\left|\int_{\mathsf{B}^c} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\int_\Delta Q_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta)\right| \leq \Pi(\mathsf{B}^c)^{1/2}\sqrt{\int \Pi(\mathrm{d}\delta,\mathrm{d}\theta)|g_k(\delta,\theta)|^2} \leq \Pi(\mathsf{B}^c)^{1/2}\frac{C_0}{1-\tilde{\lambda}}\Pi(V|\mathcal{D})^{1/2}.$$

Similarly,

$$\left|\int_{\mathsf{B}^c} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\int_\Delta \tilde{Q}_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta)\right| \leq \Pi(\mathsf{B}^c)^{1/2}\sqrt{\int \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\int_\Delta \tilde{Q}_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')|g_k(\delta',\theta)|^2} \leq \Pi(\mathsf{B}^c)^{1/2}\frac{C_0}{1-\tilde{\lambda}}\Pi(V|\mathcal{D})^{1/2},$$

using (11). Finally, since g_k is bounded on B by $C_0/(1-\tilde{\lambda})$, we have

$$\int_{\mathsf{B}} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\left|\int_\Delta Q_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta) - \int_\Delta \tilde{Q}_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')g_k(\delta',\theta)\right| \leq \frac{C_0}{1-\tilde{\lambda}}\int_{\mathsf{B}} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\left\|Q_{\theta,\mathsf{J}}(\delta,\cdot) - \tilde{Q}_{\theta,\mathsf{J}}(\delta,\cdot)\right\|_{\mathrm{tv}}.$$

It follows that

$$\left|\Pi(f) - \tilde{\Pi}(f)\right| \leq C_0\tilde{\lambda}^k + \frac{C_0\sqrt{\Pi(\mathsf{B}^c)}}{1-\tilde{\lambda}} + \frac{C_0}{1-\tilde{\lambda}}\int_{\mathsf{B}} \Pi(\mathrm{d}\delta,\mathrm{d}\theta)\left\|Q_{\theta,\mathsf{J}}(\delta,\cdot) - \tilde{Q}_{\theta,\mathsf{J}}(\delta,\cdot)\right\|_{\mathrm{tv}}.$$

The result then follows by taking $k \to \infty$. □

Appendix B. Proof of Proposition 2

Throughout C_0 denotes a generic constant. The Markov kernel of Algorithm 2 is

$$\tilde{K}((\delta,\theta);(\mathrm{d}\delta',\mathrm{d}\theta')) = K_\delta(\theta,\mathrm{d}\theta') \sum_{\mathsf{J}:\ |\mathsf{J}|=J} \binom{p}{J}^{-1} \tilde{Q}_{\theta',\mathsf{J}}(\delta,\mathrm{d}\delta'),$$

where

$$K_\delta(\theta,\mathrm{d}\theta') \stackrel{\mathrm{def}}{=} P_\delta([\theta]_\delta, \mathrm{d}[\theta']_\delta) \prod_{j:\ \delta_j=0} \mathbf{N}(0,\rho_0^{-1})(\mathrm{d}\theta'_j).$$

Recall that $V(\delta,\theta) = \sum_j \delta_j V_j(\theta_j)$. Given a selection $\mathsf{J} = \{j_1,\ldots,j_J\} \subseteq \{1,\ldots,p\}$, and $j_i \in \mathsf{J}$, we have

$$\int_\Delta \tilde{Q}_{\theta,j_i}(\delta,\mathrm{d}\delta')V(\delta',\theta) = V(\delta,\theta) + \tilde{q}_{j_i} V_{j_i}(\theta_{j_i}) - \delta_{j_i} V_{j_i}(\theta_{j_i}) \leq V(\delta,\theta) + (1-\delta_{j_i})V_{j_i}(\theta_{j_i}),$$

where $\tilde{q}_j = \tilde{q}_j(\vartheta,\theta)$. It follows that

$$\int_\Delta \tilde{Q}_{\theta,\mathsf{J}}(\delta,\mathrm{d}\delta')V(\delta',\theta) \leq V(\delta,\theta) + \sum_{i=1}^{J} V_{j_i}(\theta_{j_i}) \mathbf{1}_{\{\delta_{j_i}=0\}}. \tag{5}$$

Note that in deriving (5) we did not use any specific information about the probability $\tilde{q}_j$. In particular the kernel $Q_{\theta,\mathsf{J}}$ also satisfies (5). Using (5) we have

$$\begin{aligned}\int_{\mathbb{R}^p} K_\delta(\theta,\mathrm{d}\theta') \int_\Delta \tilde{Q}_{\theta',\mathsf{J}}(\delta,\mathrm{d}\delta')V(\delta',\theta') \\ \leq \int_{\mathbb{R}^{\|\delta\|_0}} P_\delta([\theta]_\delta,\mathrm{d}u)V_\delta(u) + \sum_{i=1}^{J} \int_{\mathbb{R}} V_{j_i}(x)\mathbf{N}(0,\rho_0^{-1})(\mathrm{d}x) \\ \leq \lambda_\delta V(\delta,\theta) + b_\delta + C_0 J,\end{aligned}$$

where the first inequality uses the fact that under K_δ, when $\delta_j = 0$ we update θ_j by drawing from $\mathbf{N}(0,\rho_0^{-1})$. With $\lambda = \max_\delta \lambda_\delta$, we conclude that

$$\int_{\Delta\times\mathbb{R}^p} \tilde{K}((\delta,\theta);(\mathrm{d}\delta',\mathrm{d}\theta'))V(\delta',\theta') \leq \lambda V(\delta,\theta) + C_0. \tag{6}$$

Furthermore, $\tilde{K}$ is phi-irreducible and aperiodic by assumption, and the level sets $\{(\delta,\theta):\ V(\delta,\theta) \leq b\}$ are petite sets for $\tilde{K}$. Therefore, by Lemma 15.2.8, and Theorem 15.0.1 of Meyn and Tweedie (2009) $\tilde{K}$ admits a unique invariant distribution $\tilde{\Pi}$, and (10) holds.

□

Appendix C. Proof of Corollary 4

Throughout the subset J is fixed. Given $\theta \in \mathbb{R}^p$, and $1 \leq j \leq p$, we recall that

$$Q_{\theta,j}(\delta,\delta') = q_j(\delta,\theta)^{\delta'_j}(1-q_j(\delta,\theta))^{1-\delta'_j}\prod_{i\neq j}\mathbf{1}_{\{\delta_i=\delta'_i\}},\quad \delta,\delta'\in\Delta,$$

and we define $\tilde{Q}_{\theta,j}$ as

$$\tilde{Q}_{\theta,j}(\delta,\delta') \stackrel{\text{def}}{=} \tilde{q}_j(\vartheta,\theta)^{\delta'_j}(1-\tilde{q}_j(\vartheta,\theta))^{1-\delta'_j}\prod_{i\neq j}\mathbf{1}_{\{\delta_i=\delta'_i\}},\quad \delta,\delta'\in\Delta,$$

where $\vartheta = \vartheta(\mathsf{J},\delta)$ is as defined in (7) in the main document, and depends on δ. As Bernoulli updates, it is easy to see that total variational distance between $Q_{\theta,j}(\delta,\cdot)$ and $\tilde{Q}_{\theta,j}(\delta,\cdot)$ is

$$\|Q_{\theta,j}(\delta,\cdot)-\tilde{Q}_{\theta,j}(\delta,\cdot)\|_{\text{tv}} = |q_j-\tilde{q}_j| \leq \min\left(\min(q_j,\tilde{q}_j), 1-\max(q_j,\tilde{q}_j)\right), \tag{7}$$

where q_j (resp. $\tilde{q}_j$) is a short for $q_j(\delta,\theta)$ (resp. $\tilde{q}_j(\vartheta,\theta)$). An analogous application of (1) then gives

$$\left\|Q_{\theta,\mathsf{J}}(\delta,\cdot)-\tilde{Q}_{\theta,\mathsf{J}}(\delta,\cdot)\right\|_{\text{tv}} \leq \sum_{k=0}^{J-1}\int_\Delta\left(Q_{\theta,j_1}\times\cdots\times Q_{\theta,j_k}\right)(\delta,\mathrm{d}\delta')\left|q_j(\delta',\theta)-\tilde{q}_j(\vartheta,\theta)\right|,$$

where for $k=0$ the product $Q_{\theta,j_{1:k}} \stackrel{\text{def}}{=} Q_{\theta,j_1}\times\cdots\times Q_{\theta,j_k}$ is the identity kernel. Note that if $(\delta,\theta)\sim\Pi(\cdot|\mathcal{D})$, and $\delta'|\theta\sim Q_{\theta,j_{1:k}}(\delta,\cdot)$, then $(\delta',\theta)\sim\Pi(\cdot|\mathcal{D})$. This implies that

$$\begin{aligned}\int_{\mathsf{B}}\Pi(\mathrm{d}\delta,\mathrm{d}\theta)\int_\Delta Q_{\theta,j_{1:k}}(\delta,\mathrm{d}\delta')\left|q_j(\delta',\theta)-\tilde{q}_j(\vartheta,\theta)\right| &\leq \Pi(\mathsf{B}^c|\mathcal{D})\\ &+\int_{\mathsf{B}}\Pi(\mathrm{d}\delta,\mathrm{d}\theta)\int_\Delta Q_{\theta,j_{1:k}}(\delta,\mathrm{d}\delta')\left|q_j(\delta',\theta)-\tilde{q}_j(\vartheta,\theta)\right|\mathbf{1}_{\mathsf{B}}(\delta',\theta).\end{aligned}$$

It then follows that

$$\begin{aligned}\left\|Q_{\theta,\mathsf{J}}(\delta,\cdot)-\tilde{Q}_{\theta,\mathsf{J}}(\delta,\cdot)\right\|_{\text{tv}} &\leq J\,\Pi(\mathsf{B}^c|\mathcal{D})\\ &+J\max_{0\leq k\leq J-1}\sup_{(\delta,\theta)\in\mathsf{B}}\int_\Delta Q_{\theta,j_{1:k}}(\delta,\mathrm{d}\delta')\left|q_j(\delta',\theta)-\tilde{q}_j(\vartheta,\theta)\right|\mathbf{1}_{\mathsf{B}}(\delta',\theta).\end{aligned}\tag{8}$$

Define $\epsilon \stackrel{\text{def}}{=} (Y-X\theta_\star)/\sigma$. Using the sub-Gaussianity of the regression error term in H2-(2), and by a union bound argument, we can choose $c>0$ depending solely on the absolute constant c_1 in H2-(2), such that

$$\mathbb{P}_\star\left(\max_{1\leq j\leq p}|\langle X_j,\epsilon\rangle| > \sqrt{cn\log(p)}\right)\leq\frac{1}{p}. \tag{9}$$

Take $(\delta,\theta)\in\mathsf{B}$, and $(\delta',\theta)\in\mathsf{B}$, and suppose that ϵ satisfies $\max_{1\leq j\leq p}\ |\langle X_j,\epsilon\rangle|\leq\sqrt{cn\log(p)}$. We show below that

$$\left|q_j(\delta',\theta)-\tilde q_j(\vartheta,\theta)\right|\leq\max\left(e^{-C_1\left(n\underline{\theta}_\star^2-(1+s_\star)\sqrt{n\log(p)}\right)},\frac{1}{\sqrt{n}}e^{-(\mathsf{u}-C(1+s_\star))\log(p)}\right).\quad(10)$$

The corollary then follows by combining (10) and (8). It remains to show (10). We consider separately the cases $\delta_{\star,j}=1$ and $\delta_{\star,j}=0$.
First, we suppose that $\delta_{\star j}=1$. We note that for all (δ,θ),

$$\begin{aligned}q_j(\delta,\theta)&=\frac{1}{1+\exp\left(\mathsf{a}+\frac{1}{2}(\rho_1-\rho_0)\theta_j^2+\ell(\theta_{\delta^{(j,0)}})-\ell(\theta_{\delta^{(j,1)}})\right)}\\&\geq\frac{1}{1+\exp\left(\mathsf{a}+\ell(\theta_{\delta^{(j,0)}})-\ell(\theta_{\delta^{(j,1)}})\right)}\geq1-\exp\left(\mathsf{a}+\ell(\theta_{\delta^{(j,0)}})-\ell(\theta_{\delta^{(j,1)}})\right).\end{aligned}$$

Using (5) in the main document, we have

$$\begin{aligned}\ell(\theta_{\delta^{(j,0)}})-\ell(\theta_{\delta^{(j,1)}})&=-\frac{\theta_j}{\sigma^2}\langle X_j,Y-X\theta_{\delta^{(j,0)}}\rangle+\frac{\theta_j^2n}{2\sigma^2}\\&=-\frac{\theta_j}{\sigma}\langle X_j,\epsilon\rangle-\frac{\theta_j\theta_{\star j}n}{\sigma^2}+\frac{\theta_j}{\sigma^2}\sum_{k=1}^p(\delta_k^{(j,0)}\theta_k-\delta_{\star k}^{(j,0)}\theta_{\star k})\langle X_k,X_j\rangle+\frac{\theta_j^2n}{2\sigma^2}.\end{aligned}\quad(11)$$

For $(\delta,\theta)\in\mathsf{B}$, and under H2, we have

$$\ell(\theta_{\delta^{(j,0)}})-\ell(\theta_{\delta^{(j,1)}})\leq-\frac{\theta_j\theta_{\star j}n}{\sigma^2}+\frac{\theta_j^2n}{2\sigma^2}+C\left(\sqrt{n\log(p)}+\|\delta\|_0\|\theta-\theta_\star\|_\infty\sqrt{n\log(p)}\right),$$

and

$$-\theta_j\theta_{\star j}n+\frac{\theta_j^2n}{2}\leq-\frac{n\theta_{\star j}^2}{2}+C\left(\log(p)+\sqrt{n\log(p)}\right).$$

Combining the last two inequalities yields,

$$\ell(\theta_{\delta^{(j,0)}})-\ell(\theta_{\delta^{(j,1)}})\leq-\frac{n\theta_{\star j}^2}{2}+C\left(\sqrt{n\log(p)}+s\log(p)\right),$$

for some constant C_1. We conclude that for $n\geq s^2\log(p)$,

$$q_j(\delta,\theta)\geq1-e^{\mathsf{a}-C_1\left(n\underline{\theta}_\star^2-\sqrt{n\log(p)}\right)},$$

for some constant C_1. Similarly, for $(\delta,\theta)\in\mathsf{B}$, and ϑ as in (7) of the main document

$$\tilde q_j(\vartheta,\theta)\geq1-\exp\left(\mathsf{a}-\frac{\theta_j}{\sigma^2}\langle X_j,Y-X\theta_\vartheta\rangle\right).$$

As in (11),

$$-\frac{\theta_j}{\sigma^2}\left\langle X_j, Y - X\theta_\vartheta\right\rangle = -\frac{\theta_j}{\sigma}\left\langle X_j, \epsilon\right\rangle - \frac{\theta_j\theta_{\star j}n}{\sigma^2} + \frac{\theta_j}{\sigma^2}\sum_{k=1}^{p}(\vartheta_k\theta_k - \delta_{\star k}^{(j,0)}\theta_{\star k})\left\langle X_k, X_j\right\rangle$$
$$\leq -\frac{n\theta_{\star j}^2}{2} + C\left(\sqrt{n\log(p)} + (s-s_\star)\log(p) + s_\star\sqrt{n\log(p)}\right)$$
$$-\frac{n\theta_{\star j}^2}{2} + C(1+s_\star)\sqrt{n\log(p)}$$

We conclude that

$$\tilde{q}_j(\vartheta,\theta) \geq 1 - e^{\mathsf{a} - C\left(n\underline{\theta}_\star^2 - (1+s_\star)\sqrt{n\log(p)}\right)},$$

and since $e^{\mathsf{a}} = \sqrt{n}p^{\mathsf{u}}/\sigma$, we get

$$|q_j(\delta,\theta) - \tilde{q}_j(\vartheta,\theta)| \leq \frac{\sqrt{n}p^{\mathsf{u}}}{\sigma}e^{-C\left(n\underline{\theta}_\star^2 - (1+s_\star)\sqrt{n\log(p)}\right)},$$

for some constant C.

Suppose now that $\delta_{\star j} = 0$. The we use

$$q_j(\delta,\theta) \leq \exp\left(-\mathsf{a} + \frac{\rho_0\theta_j^2}{2} - \left(\ell(\theta_{\delta^{(j,0)}}) - \ell(\theta_{\delta^{(j,1)}})\right)\right).$$

Starting from (11), we obtain

$$\ell(\theta_{\delta^{(j,0)}}) - \ell(\theta_{\delta^{(j,1)}}) \leq C\log(p).$$

It follows that

$$q_j(\delta,\theta) \leq \frac{\sigma e^{C\log(p)}}{\sqrt{n}p^{\mathsf{u}}}.$$

Similarly,

$$\tilde{q}_j(\vartheta,\theta) \leq \exp\left(-\mathsf{a} + \frac{\rho_0^2\theta_j^2}{2} + \frac{n\theta_j^2}{2\sigma^2} + \frac{\theta_j}{\sigma^2}\left\langle X_j, Y - X\theta_\vartheta\right\rangle\right),$$

and

$$\left|\frac{\theta_j}{\sigma^2}\left\langle X_j, Y - X\theta_\vartheta\right\rangle\right| \leq C(1+s_\star)\log(p),$$

which leads to

$$|q_j(\delta,\theta) - \tilde{q}_j(\vartheta,\theta)| \leq \frac{1}{\sqrt{n}p^{\mathsf{u}}}e^{C(1+s_\star)\log(p)}.$$

Combining the two cases proves (10).

Appendix D. Proof of Theorem 5

We recall that $\mathbb{P}$ and $\mathbb{E}$ denote the probability measure and expectation operator of the Markov chains defined by Algorithms 1 and 2 (more specifically their coupling distribution as constructed below), and $\mathbb{P}_\star$ and $\mathbb{E}_\star$ denote the probability measure and expectation operator of the data Y as assumed in H2.

Throughout we will use C to denote a generic constant that depends only on the constants appearing in H2 ($\sigma^2, \|\theta_\star\|_\infty$, c_0, c_1 and c_2). The actual value of C may vary from one appearance to the next.

We use a similar argument as in De Sa et al. (2016). Let $\{\delta^{(k)},\ k \geq 0\}$ denote the δ-marginal chain of Algorithm 2, and let $\{\check{\delta}^{(k)},\ k \geq 0\}$ be the δ-marginal chain of Algorithm 1. These processes are also Markov chains because in both cases we have taken $P_\delta = \tilde{P}_\delta$ to be an exact draw from the posterior conditional distribution of θ given δ. We construct a coupling of $\{\delta^{(k)},\ k \geq 0\}$ and the stationary version of $\{\check{\delta}^{(k)},\ k \geq 0\}$ as follows. First take $\delta^{(0)}$ as the null model, and draw $\check{\delta}^{(0)} \sim \Pi(\cdot|\mathcal{D})$, the marginal distribution of δ in (2). For each $k \geq 0$, given $(\delta^{(k)}, \check{\delta}^{(k)})$, we do the following.

(1) Given $\delta^{(k)}, \check{\delta}^{(k)}$, we independently draw $\theta^{(k)} \sim \Pi(\cdot|\delta^{(k)}, \mathcal{D})$, $\check{\theta}^{(k)} \sim \Pi(\cdot|\check{\delta}^{(k)}, \mathcal{D})$, and we select a random subset $\mathsf{J}^{(k)} = \{\mathsf{J}_1^{(k)}, \ldots, \mathsf{J}_J^{(k)}\}$ of size J from $\{1, \ldots, p\}$.

(2) We define $\vartheta \in \Delta$ as $\vartheta_i = 0$ if $i \in \mathsf{J}^{(k)}$, and $\vartheta_i = \delta_i^{(k)}$ otherwise. We also define $\vartheta^{(0)} = \delta^{(k)}$, and $\check{\vartheta}^{(0)} = \check{\delta}^{(k)}$. For each $r \in \{1, \ldots, J\}$, given $\mathsf{J}_r^{(k)} = j$, we then do the following.
   (a) We draw $(d_r^{(k)}, \check{d}_r^{(k)})$ from the maximal coupling of $\mathsf{Ber}(\tilde{q}_j(\vartheta, \theta^{(k)}))$ and $\mathsf{Ber}(q_j(\check{\vartheta}^{(r-1)}, \check{\theta}^{(k)}))$, where q_j and $\tilde{q}_j$ are given by (4) and (6) respectively.
   (b) We set $\vartheta_j^{(r)} = d_r^{(k)}$, $\check{\vartheta}_j^{(r)} = \check{d}_r^{(k)}$, and $\vartheta_i^{(r)} = \vartheta_i^{(r-1)}$, $\check{\vartheta}_i^{(r)} = \check{\vartheta}_i^{(r-1)}$, for $i \neq j$.

(3) Set $\delta^{(k+1)} = \vartheta^{(J)}$, and $\check{\delta}^{(k+1)} = \check{\vartheta}^{(J)}$.

By construction, the marginal chain $\{\delta^{(k)}\ k \geq 0\}$ (resp. $\{\check{\delta}^{(k)}\ k \geq 0\}$) from the above construction is the asynchronous sampler from Algorithm 2 (resp. a stationary version of Algorithm 1). By the coupling inequality

$$\mathbb{E}_\star \left[\max_{j\colon \delta_{\star j}=1} \left|\mathbb{P}(\delta_j^{(k)} = 1) - \Pi(\delta_j = 1|\mathcal{D})\right| \right] \leq \mathbb{E}_\star \left[\max_{j\colon \delta_{\star j}=1} \mathbb{P}(\delta_j^{(k)} \neq \check{\delta}_j^{(k)}) \right]. \tag{12}$$

Hence the main part of the proof consists in bounding the right-hand side of the last display. We do this in paragraph (e). Paragraphs (a)-(d) collect some needed implications of H2.

`(a) Restricted eigenvalues`. Given $s > 0$, Let $\delta \in \Delta$ be such that $0 < \|\delta\|_0 \leq s$, and let $u \in \mathbb{R}^{\|\delta\|_0}$. Using $|\langle X_i, X_j\rangle| \leq \sqrt{c_0 n \log(p)}$ from H2-(1), and $\|X_j\|_2 = \sqrt{n}$, we have

$$u'(X_\delta' X_\delta)u \geq n\|u\|_2^2 - \sqrt{c_0 n \log(p)} \sum_{i\neq j} |u_i u_j| \geq \left(n - s\sqrt{c_0 n\log(p)}\right) \|u\|_2^2.$$

We conclude that if the sample size satisfies $n \geq 4c_0 s^2 \log(p)$, then

$$\lambda_{\min}\left(X_\delta' X_\delta\right) \geq \frac{n}{2}, \quad \text{for all } \delta \in \Delta, \quad \text{s. t.} \quad 0 < \|\delta\|_0 \leq s, \tag{13}$$

where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A.

`(b) Implications of the sub-Gaussian regression errors.` For $\delta \in \Delta$, we define

$$L_\delta \stackrel{\text{def}}{=} I_n + \frac{1}{\sigma^2} X_\delta X_\delta'.$$

We convene that $L_\delta = I_n$, for $\delta = 0$. Clearly, $\|L_\delta^{-1}\|_2 \leq 1$. Given $s \geq 0$ (here we allow s to be 0), and for some constant $c > 0$, we set

$$\mathcal{E}_s \stackrel{\text{def}}{=} \left\{ y \in \mathbb{R}^n : \max_{1\leq j\leq p} \max_{\delta:\ \|\delta\|_0 \leq s} \sigma^{-1} \left|\left\langle L_\delta^{-1} X_j, y - X\theta_\star\right\rangle\right| \leq \sqrt{c(1+s)n\log(p)} \right\}.$$

Using the sub-Gaussianity of the regression error term in H2-(2), and by a union bound argument, we can choose $c > 0$ depending solely on the absolute constant c_1 in H2-(2), such that for all $s \geq 0$,

$$\mathbb{P}_\star\left(Y \notin \mathcal{E}_s\right) \leq p \sum_{j=0}^{s} \binom{p}{j} c_1 \exp\left(-\frac{c(1+s)n\log(p)}{2n}\right) \leq \frac{c_1 p^{s+1}}{p^{\frac{c(1+s)}{2}}} \leq \frac{1}{p}, \tag{14}$$

where we use the fact that $\sum_{j=0}^{s}\binom{p}{j} \leq 2p^s$. Throughout the proof, whenever we use the event $\mathcal{E}_s$, the constant c is always taken as above.

`(c) Sparse MCMC output.` It will be important in the proof to guarantee that the Markov chain $\{\delta^{(k)},\ k \geq 1\}$ remains in the set $\Delta_s \stackrel{\text{def}}{=} \{\delta \in \Delta :\ \|\delta\|_0 \leq s\}$ for some small value of s. The following result may be improved, but will serve the purpose. Let

$$s_1 \stackrel{\text{def}}{=} s_\star + 2\log(p).$$

We show in Lemma 1 that under the sample size condition (19), and $\mathtt{u}$ taken large enough as in (19), it holds

$$\mathbf{1}_{\mathcal{E}_0}(Y) \max_{k\geq 0} \mathbb{P}\left(\|\delta^{(k)}\|_0 > s_1\right) \leq \frac{1}{p}. \tag{15}$$

`(d) Posterior contraction`. We show below that the posterior distribution $\Pi(\cdot|\mathcal{D})$ puts most probability mass on sparse super-sets of $\delta_\star$. More precisely, by Lemma 3 we can find constants C_1, C_2 that depends only on the constants appearing in H2 ($\sigma^2, \|\theta_\star\|_\infty$, c_0, c_1 and c_2) such that for n, p such that $n \geq C_1(1+s_\star^3)\log(p)$, it holds

$$\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\Pi(\mathcal{C}|\mathcal{D})\right] \geq 1 - \frac{2}{p},$$

where

$$\mathcal{C} \stackrel{\text{def}}{=} \{\delta \in \Delta :\ \delta \supseteq \delta_\star, \ \text{ and } \ \|\delta\|_0 \leq C_2(1+s_\star)\}\,.$$

We set

$$s_2 \stackrel{\text{def}}{=} C_2(1+s_\star).$$

Furthermore the linear regression setting implies that the conditional posterior distribution of $\theta|\delta$ is given by

$$[\theta]_{\delta^c} \mid \delta \stackrel{i.i.d.}{\sim} \mathbf{N}(0, \rho_0^{-1}), \quad \text{and} \quad [\theta]_\delta \mid \delta \sim \mathbf{N}\left(\hat{\theta}_\delta, \sigma^2\left(\sigma^2\rho_1 I_{\|\delta\|_0} + X_\delta' X_\delta\right)^{-1}\right), \tag{16}$$

where

$$\hat{\theta}_\delta \stackrel{\text{def}}{=} \text{Argmax}_{u\in\mathbb{R}^{\|\delta\|_0}}\left[-\frac{1}{2\sigma^2}\|y - X_\delta u\|_2^2 - \frac{\rho_1}{2}\|u\|_2^2\right] = \left(X_\delta' X_\delta + \rho_1\sigma^2 I_{\|\delta\|_0}\right)^{-1} X_\delta' y.$$

Therefore, if for some $M>0$ we set

$$\mathsf{B}_\delta \stackrel{\text{def}}{=} \left\{\theta \in \mathbb{R}^p :\ \|\theta - \theta_\delta\|_\infty \leq \sqrt{\frac{M\log(p)}{\rho_0}} \quad \text{and} \quad \|\theta_\delta - \hat{\theta}_\delta\|_\infty \leq \sqrt{\frac{M\sigma\log(p)}{n}}\right\},$$

then, provided that $n \geq Cs^2\log(p)$, for some constant C, by the restricted eigenvalue bound in (13), and by Gaussian tail bounds and a union bound argument, for all $\delta \in \Delta_s$, we have

$$\Pi\left(\mathsf{B}_\delta^c|\delta, \mathcal{D}\right) \leq \frac{4}{p^{-1+M/2}} \leq \frac{1}{p}, \tag{17}$$

by taking $M>2$ appropriately.

`(e) Main arguments of the proof`. With s_1 as in Paragraph (c) and s_2 as in Paragraph (d), we set

$$s \stackrel{\text{def}}{=} \max(s_1, s_2).$$

Fix $Y \in \mathcal{E}_{s_\star}$, and fix some arbitrary component j such that $\delta_{\star j} = 1$. We first note that $\delta_j^{(k+1)} \neq \check{\delta}_j^{(k+1)}$ if and only if $j \in \mathsf{J}^{(k)}$, and the corresponding Bernoulli's $(d_r^{(k)}, \check{d}_r^{(k)})$

are different, or $\delta_j^{(k)} \neq \check{\delta}_j^{(k)}$, and $j \notin \mathsf{J}^{(k)}$. We write this as

$$\begin{aligned}
&\mathbb{P}\left[\delta_j^{(k+1)} \neq \check{\delta}_j^{(k+1)} | \delta^{(k)}, \check{\delta}^{(k)}\right] \\
&\qquad = \mathbf{1}_{\left\{\delta_j^{(k)} \neq \check{\delta}_j^{(k)}\right\}} \left(1 - \frac{J}{p}\right) + \sum_{r=1}^{J} \mathbb{P}\left[\mathsf{J}_r^{(k)} = j, d_r^{(k)} \neq \check{d}_r^{(k)} | \delta^{(k)}, \check{\delta}^{(k)}\right] \\
&\qquad = \mathbf{1}_{\left\{\delta_j^{(k)} \neq \check{\delta}_j^{(k)}\right\}} \left(1 - \frac{J}{p}\right) + \frac{1}{p}\sum_{r=1}^{J} \mathbb{P}\left[d_r^{(k)} \neq \check{d}_r^{(k)} | \mathsf{J}_r^{(k)} = j, \delta^{(k)}, \check{\delta}^{(k)}\right],
\end{aligned}$$

where we use the fact that $\mathbb{P}(\mathsf{J}_r^{(k)} = j | \delta^{(k)}, \check{\delta}^{(k)}) = 1/p$. With s_1, s_2 as above, we introduce the set $\mathbb{T} \stackrel{\text{def}}{=} \Delta_{s_1} \times \mathcal{C}_{s_2}$ where,

$$\Delta_{s_1} \stackrel{\text{def}}{=} \left\{\delta \in \Delta :\ \|\delta\|_0 \leq s_1\right\}, \quad \text{and} \quad \mathcal{C}_{s_2} \stackrel{\text{def}}{=} \left\{\delta \in \Delta :\ \delta \supseteq \delta_\star,\ \|\delta\|_0 \leq s_2\right\}.$$

It follows that

$$\begin{aligned}
&\mathbb{P}\left[\delta_j^{(k+1)} \neq \check{\delta}_j^{(k+1)} | \delta^{(k)}, \check{\delta}^{(k)}\right] \\
&\quad \leq \mathbf{1}_{\left\{\delta_j^{(k)} \neq \check{\delta}_j^{(k)}\right\}} \left(1 - \frac{J}{p}\right) + \mathbf{1}_{\mathbb{T}}(\delta^{(k)}, \check{\delta}^{(k)}) \frac{1}{p}\sum_{r=1}^{J} \mathbb{P}\left[d_r^{(k)} \neq \check{d}_r^{(k)} | \mathsf{J}_r^{(k)} = j, \delta^{(k)}, \check{\delta}^{(k)}\right] \\
&\qquad\qquad + \frac{J}{p} \mathbf{1}_{\mathbb{T}^c}(\delta^{(k)}, \check{\delta}^{(k)}).
\end{aligned}$$

Let us set

$$A^{(k)} \stackrel{\text{def}}{=} \mathbb{P}\left((\delta^{(k)}, \check{\delta}^{(k)}) \notin \mathbb{T}\right),\ \ \mathcal{I}_{r,j}^{(k)}(\theta, \check{\theta}) \stackrel{\text{def}}{=} \mathbb{P}\left(d_r^{(k)} \neq \check{d}_r^{(k)} | \mathsf{J}_r^{(k)} = j, \delta^{(k)}, \check{\delta}^{(k)}, \theta, \check{\theta}\right).$$

Taking expectation on both sides of the last inequality, we get

$$\begin{aligned}
\mathbb{P}\left(\delta_j^{(k+1)} \neq \check{\delta}_j^{(k+1)}\right) \quad &\leq \quad \left(1 - \frac{J}{p}\right) \mathbb{P}\left(\delta_j^{(k)} \neq \check{\delta}_j^{(k)}\right) + \frac{JA^{(k)}}{p} \\
&\quad + \frac{1}{p}\sum_{r=1}^{J} \mathbb{E}\left[\mathbf{1}_{\mathbb{T}}(\delta^{(k)}, \check{\delta}^{(k)}) \mathbb{P}\left(d_r \neq \check{d}_r | \mathsf{J}_r^{(k)} = j, \delta^{(k)}, \check{\delta}^{(k)}\right)\right], \\
&= \quad \left(1 - \frac{J}{p}\right) \mathbb{P}\left(\delta_j^{(k)} \neq \check{\delta}_j^{(k)}\right) + \frac{JA^{(k)}}{p} \\
&\quad + \frac{1}{p}\sum_{r=1}^{J} \mathbb{E}\left[\mathbf{1}_{\mathbb{T}}(\delta^{(k)}, \check{\delta}^{(k)}) \int \mathcal{I}_{r,j}^{(k)}(\theta, \check{\theta})\, \Pi(\mathrm{d}\theta | \delta^{(k)}, \mathcal{D})\, \Pi(\mathrm{d}\check{\theta} | \check{\delta}^{(k)}, \mathcal{D})\right].
\end{aligned} \tag{18}$$

We establish the following claim below

$$\mathbf{1}_{\mathbb{T}}(\delta^{(k)},\check{\delta}^{(k)})\int \mathcal{I}_{r,j}^{(k)}(\theta,\check{\theta})\,\Pi(\mathrm{d}\theta|\delta^{(k)},\mathcal{D})\,\Pi(\mathrm{d}\check{\theta}|\check{\delta}^{(k)},\mathcal{D})$$
$$\leq \left(e^{-C\sqrt{n}\underline{\theta}_\star}+\frac{1}{p}\right)+\frac{7}{10}\mathbf{1}_{\{\delta_j^{(k)}\neq\check{\delta}_j^{(k)}\}}. \quad (19)$$

Using (19) in (18), we obtain

$$\begin{aligned}\mathbb{P}\left(\delta_j^{(k+1)}\neq\check{\delta}_j^{(k+1)}\right) &\leq \left(1-\frac{J}{p}\right)\mathbb{P}\left(\delta_j^{(k)}\neq\check{\delta}_j^{(k)}\right)+\frac{JA^{(k)}}{p}\\ &\quad+\frac{J}{p}\left(e^{-C\sqrt{n}\underline{\theta}_\star}+\frac{1}{p}\right)+\frac{7}{10}\frac{J}{p}\mathbb{P}\left(\delta_j^{(k)}\neq\check{\delta}_j^{(k)}\right)\\ &\leq \left(1-\frac{3}{10}\frac{J}{p}\right)\mathbb{P}\left(\delta_j^{(k)}\neq\check{\delta}_j^{(k)}\right)+\frac{J}{p}\left(A^{(k)}+e^{-C\sqrt{n}\underline{\theta}_\star}+\frac{1}{p}\right). \quad (20)\end{aligned}$$

Iterating (20) yields

$$\begin{aligned}\max_{j:\,\delta_{\star j}=1}\mathbb{P}\left(\delta_j^{(k)}\neq\check{\delta}_j^{(k)}\right) &\leq \left(1-\frac{3}{10}\frac{J}{p}\right)^k+\frac{10}{3}\left(e^{-C\sqrt{n}\underline{\theta}_\star}+\frac{1}{p}\right)\\ &\quad+\frac{J}{p}\sum_{t=0}^{k-1}\left(1-\frac{3}{10}\frac{J}{p}\right)^t A^{(k-t)}. \quad (21)\end{aligned}$$

Recall that

$$A^{(k)}\stackrel{\text{def}}{=}\mathbb{P}\left((\delta^{(k)},\check{\delta}^{(k)})\notin\mathbb{T}\right)\leq\mathbb{P}(\|\delta^{(k)}\|_0>s_1)+\Pi(\mathcal{C}_{s_2}^c|\mathcal{D}),$$

where $\mathcal{C}_{s_2}^c\stackrel{\text{def}}{=}\Delta\setminus\mathcal{C}_{s_2}$. By Lemma 1 and Lemma 3 below, we have

$$\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\max_{k\geq 0}\mathbb{P}(\|\delta^{(k)}\|_0>s_1)\leq\frac{1}{p},\quad\text{and}\quad \mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\Pi(\mathcal{C}_{s_2}^c|\mathcal{D})\right]\leq\frac{1}{p}.$$

Taking the expectation over the data Y in (21), and using the last display and (12), we deduce that

$$\begin{aligned}\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\max_{j:\,\delta_{\star j}=1}\left|\mathbb{P}(\delta_j^{(k)}=1)-\Pi(\delta_j=1|\mathcal{D})\right|\right] &\leq\left(1-\frac{3}{10}\frac{J}{p}\right)^k\\ &\quad+\frac{10}{3}\left(e^{-C\sqrt{n}\underline{\theta}_\star}+\frac{1}{p}+\frac{2}{p}\right)\\ &\leq\left(1-\frac{3}{10}\frac{J}{p}\right)^k+10\left(e^{-C\sqrt{n}\underline{\theta}_\star}+\frac{1}{p}\right). \quad (22)\end{aligned}$$

It remains only to establish the claim (19).

`Proof of Claim (19).` We consider two cases.

<u>`Case 1:` $\delta_j^{(k)} \neq \check{\delta}_j^{(k)}$.</u> Since $\check{\delta}^{(k)} \in \mathcal{C}_{s_1}$ (which implies that $\check{\delta}_j^{(k)} = 1$), we must then have $\delta_j^{(k)} = 0$, and $\check{\delta}_j^{(k)} = 1$. Set

$$\mathbb{S} \stackrel{\text{def}}{=} \left\{(\theta, \check{\theta}) \in \mathbb{R}^p \times \mathbb{R}^p : \; \theta \in \mathsf{B}_{\delta^{(k)}}, \check{\theta} \in \mathsf{B}_{\check{\delta}^{(k)}}, \quad \text{and} \quad \sqrt{\frac{1}{100\rho_0}} \leq \mathsf{sign}(\theta_{\star j})\theta_j \leq \sqrt{\frac{4}{\rho_0}}\right\}.$$

It follows from (17) and the fact that $\theta_j^{(k)} | \{\delta_j^{(k)} = 0\} \sim \mathbf{N}(0, \rho_0^{-1})$ that for $(\delta^{(k)}, \check{\delta}^{(k)}) \in \mathbb{T}$,

$$\mathbb{P}\left((\theta^{(k)}, \check{\theta}^{(k)}) \notin \mathbb{S} | \delta^{(k)}, \check{\delta}^{(k)}\right) \leq \frac{4}{p^{M/2}} + \frac{3}{5} \leq \frac{7}{10}. \tag{23}$$

First we note that for $(\theta, \check{\theta}) \in \mathbb{S}$, $|\theta_j| \leq 2/\sqrt{\rho_0} \leq Cn^{-1/2}$. Whereas for $i \neq j$ and $\delta_i^{(k)} = 0$, we have $|\theta_i| \leq \sqrt{M \log(p)/n}$, and if $\delta_i^{(k)} = 1$, using (17), and Lemma 2-(1),

$$|\theta_i| = |\theta_i - \hat{\theta}_i| + |\hat{\theta}_i - \theta_{\star i}| + |\theta_{\star i}| \leq C\sqrt{\frac{\log(p)}{n}} + C\sqrt{\frac{\log(p)}{n}} + \|\theta_\star\|_\infty \leq C,$$

under the sample size condition (19). Using the expression $\tilde{q}_j$ in (6), and since $\rho_0 \geq \rho_1$, and ignoring the nonpositive quadratic term, we have

$$1 - \tilde{q}_j(\vartheta, \theta) \leq \exp\left(\mathsf{a} - \frac{\theta_j}{\sigma^2}\langle X_j, y - X\theta_\vartheta\rangle\right).$$

We write

$$\langle X_j, y - X\theta_\vartheta\rangle = [\langle X_j, X\theta_{\delta^{(k)}}\rangle - \langle X_j, X\theta_\vartheta\rangle] + \langle X_j, y - X\theta_{\delta^{(k)}}\rangle.$$

Since $|\theta_i| \leq C$, for $\delta_i^{(k)} = 1$, we have

$$|\langle X_j, X\theta_\vartheta\rangle - \langle X_j, X\theta_{\delta^{(k)}}\rangle| = \left|\sum_{r \in \mathsf{J}^{(k)} : r \neq j, \; \delta_r^{(k)} = 1} \langle X_j, X_r\rangle \theta_r\right| \leq C \min(J, s_1)\sqrt{n \log(p)}.$$

We can rewrite the last display as

$$|\langle X_j, y - X\theta_\vartheta\rangle - \langle X_j, y - X\theta_{\delta^{(k)}}\rangle| \leq C \min(J, s_1)\sqrt{n \log(p)}.$$

We further expand the term $\langle X_j, y - X\theta_{\delta^{(k)}}\rangle$ as

$$\langle X_j, y - X\theta_{\delta^{(k)}}\rangle = \langle X_j, y - X\theta_\star\rangle + \langle X_j, X\theta_\star - X_{\delta^{(k)}}[\theta_\star]_{\delta^{(k)}}\rangle + \langle X_j, X_{\delta^{(k)}}([\theta_\star]_{\delta^{(k)}} - [\theta]_{\delta^{(k)}})\rangle.$$

Note that

$$\langle X_j, X\theta_\star - X_\delta[\theta_\star]_{\delta^{(k)}}\rangle = n\theta_{\star j} + \sum_{r:\; \delta_{\star r} = 1, \delta_r^{(k)} = 0} \theta_{\star r}\langle X_j, X_r\rangle.$$

For $\theta \in \mathsf{B}_{\delta^{(k)}}$,

$$
\begin{aligned}
|\langle X_j, X_{\delta^{(k)}}([\theta_\star]_{\delta^{(k)}} - [\theta]_{\delta^{(k)}})\rangle| &\le s_1\sqrt{c_0 n\log(p)}\|[\theta_\star]_{\delta^{(k)}} - [\theta]_{\delta^{(k)}}\|_\infty \\
&\le Cs_1\sqrt{n\log(p)}\left(\|[\theta_\star]_{\delta^{(k)}} - \hat{\theta}_{\delta^{(k)}}\|_\infty + \sqrt{\frac{\log(p)}{n}}\right) \\
&\le C(1+m(\delta^{(k)}))s_1\log(p),
\end{aligned}
$$

where the last inequality uses Lemma 2. Using this, and since $\delta_{\star j} = 1$, for $\theta \in \mathsf{B}_{\delta^{(k)}}$ we have

$$
\begin{aligned}
|\langle X_j, y - X\theta_\vartheta\rangle - n\theta_{\star j}| &\le C\min(s_1, J)\sqrt{n\log(p)} + |\langle X_j, y - X\theta_\star\rangle| \\
&\quad + \sum_{r:\ \delta_{\star r}=1, \delta_r^{(k)}=0} |\theta_{\star r}\langle X_j, X_r\rangle| + C(1+m(\delta^{(k)}))s_1\log(p) \\
&\le C\left(s_\star + \min(J, s_1)\right)\sqrt{n\log(p)}, \qquad (24)
\end{aligned}
$$

using the sample size condition $n \ge s_1^2\log(p)$. Since $|\theta_j| \le Cn^{-1/2}$, we conclude that

$$
|\theta_j||\langle X_j, y - X\theta_\vartheta\rangle - n\theta_{\star j}| \le C\left(s_\star + \min(J, s_1)\right)\sqrt{\log(p)}.
$$

It follows that for $(\theta, \check{\theta}) \in \mathbb{S}$,

$$
\begin{aligned}
\frac{\theta_j}{\sigma^2}\langle X_j, y - X\theta_\vartheta\rangle &\ge \frac{n\theta_{\star j}\theta_j}{\sigma^2} - C\left(s_\star + \min(J, s_1)\right)\sqrt{\log(p)} \\
&\ge \frac{|\theta_{\star j}|\sqrt{n}}{10\sigma^2} - C\left(s_\star + \min(J, s_1)\right)\sqrt{\log(p)} \ge \frac{|\theta_{\star j}|\sqrt{n}}{20\sigma^2} - CJ\sqrt{\log(p)},
\end{aligned}
$$

under the sample size condition (19). Hence, since $\mathsf{a} = \mathsf{u}\log(p) + \log(\rho_0)/2$, for $(\theta, \check{\theta}) \in \mathbb{S}$

$$
1 - \tilde{q}_j(\vartheta, \theta) \le \exp\left(\mathsf{u}\log(p) + \frac{1}{2}\log\left(\frac{n}{\sigma^2}\right) - \frac{\sqrt{n}\underline{\theta}_\star}{20\sigma}\right) \le \exp\left(-C_1\sqrt{n}\underline{\theta}_\star + C_2 J\sqrt{\log(p)}\right). \qquad (25)
$$

We handle $1 - q_j(\check{\vartheta}^{(r-1)}, \theta)$ similarly: since $\rho_0 \ge \rho_1$,

$$
1 - q_j(\check{\vartheta}^{(r-1)}, \check{\theta}) \le \exp\left(\mathsf{a} + \frac{n\check{\theta}_j^2}{2\sigma^2} - \frac{\check{\theta}_j}{\sigma^2}\left\langle X_j, y - X\check{\theta}_{(\check{\vartheta}^{(r-1)})^{(j,0)}}\right\rangle\right).
$$

The inequality (24) remains valid when applied to $\check{\theta}$ and $(\check{\vartheta}^{(r-1)})^{(j,0)}$ (but with $\min(J, s_1)$ replaced by J), and yields

$$
\left|\left\langle X_j, y - X\check{\theta}_{(\check{\vartheta}^{(r-1)})^{(j,0)}}\right\rangle - n\theta_{\star j}\right| \le C\left(s_\star + J\right)\sqrt{n\log(p)},
$$

leading to

$$-\frac{\check{\theta}_j}{\sigma^2}\left\langle X_j, y - X\check{\theta}_{(\check{\vartheta}^{(r-1)})^{(j,0)}}\right\rangle$$
$$\leq -\frac{n}{\sigma^2}\theta_{\star j}^2 + \frac{n\theta_{\star j}}{\sigma^2}|\check{\theta}_j - \theta_{\star j}| + C\left(s_\star + J\right)\sqrt{n\log(p)}$$
$$\leq \frac{n}{\sigma^2}\theta_{\star j}^2 + C\left(s_\star + J\right)\sqrt{n\log(p)},$$

where we use Lemma 2 to derive the bound $|\check{\theta}_j - \theta_{\star j}| \leq \left(|\check{\theta}_j - \hat{\theta}_j| + |\hat{\theta}_j - \theta_{\star j}|\right) \leq C\sqrt{n/\log(p)}$. The same bound implies that

$$\frac{n\check{\theta}_j^2}{2\sigma^2} = \frac{n\theta_{\star j}^2}{2\sigma^2} + \frac{n(\check{\theta}_j^2 - \theta_{\star j}^2)}{2\sigma^2} \leq \frac{n\theta_{\star j}^2}{2\sigma^2} + C\sqrt{n\log(p)}.$$

We conclude that

$$\frac{n\check{\theta}_j^2}{2\sigma^2} - \frac{\check{\theta}_j}{\sigma^2}\left\langle X_j, y - X\check{\theta}_{(\check{\vartheta}^{(r-1)})^{(j,0)}}\right\rangle$$
$$\leq -\frac{n\theta_{\star j}^2}{2\sigma^2} + C\left(s_\star + J\right)\sqrt{n\log(p)} \leq -\frac{n\underline{\theta_\star}^2}{4\sigma^2} + CJ\sqrt{n\log(p)},$$

under the sample size condition (19). Hence

$$1 - q_j(\check{\vartheta}^{(r-1)}, \check{\theta}) \leq \exp\left(\mathsf{a} - \frac{n\underline{\theta_\star}^2}{4\sigma^2}\right) \leq \exp\left(-\frac{n\underline{\theta_\star}^2}{8\sigma^2} + CJ\sqrt{n\log(p)}\right)$$
$$\leq \exp\left(-C_1\sqrt{n}\underline{\theta_\star} + C_2 J\sqrt{\log(p)}\right), \quad (26)$$

using again the sample size condition (19). Since the Bernoulli random variables $d_r^{(k)}$ and $\check{d}_r^{(k)}$ are maximally coupled, (25) and (26) imply that for $(\delta^{(k)}, \check{\delta}^{(k)}) \in \mathbb{T}$, and $\delta_j^{(k)} \neq \check{\delta}_k^{(k)}$,

$$\int \mathcal{I}_{r,j}^{(k)}(\theta, \check{\theta})\,\Pi(\mathrm{d}\theta|\delta^{(k)}, \mathcal{D})\,\Pi(\mathrm{d}\check{\theta}|\check{\delta}^{(k)}, \mathcal{D}) \leq \exp\left(-C_1\sqrt{n}\underline{\theta_\star} + C_2 J\sqrt{\log(p)}\right) + \frac{7}{10}. \quad (27)$$

<u>`Case 2`: $\delta_j^{(k)} = \check{\delta}_j^{(k)}$.</u> Since $\check{\delta}^{(k)} \in \mathcal{C}_{s_1}$, we must then have $\delta_j^{(k)} = \check{\delta}_j^{(k)} = 1$. Here we define the set $\mathbb{S}$ as

$$\mathbb{S} \stackrel{\text{def}}{=} \left\{(\theta, \check{\theta}) \in \mathbb{R}^p \times \mathbb{R}^p :\ \theta \in \mathsf{B}_{\delta^{(k)}}, \check{\theta} \in \mathsf{B}_{\check{\delta}^{(k)}}\right\}.$$

It follows from (17) that for $(\delta^{(k)}, \check{\delta}^{(k)}) \in \mathbb{T}$,

$$\mathbb{P}\left((\theta^{(k)}, \check{\theta}^{(k)}) \notin \mathbb{S}|\delta^{(k)}, \check{\delta}^{(k)}\right) \leq \frac{4}{p^{1-M/2}}. \quad (28)$$

For $(\delta^{(k)}, \check{\delta}^{(k)}) \in \mathbb{T}$, and $(\theta, \check{\theta}) \in \mathbb{S}$, the calculations on $1 - q_j(\check{\vartheta}^{(r-1)}, \check{\theta})$ remain valid, and we have

$$1 - q_j(\check{\vartheta}^{(r-1)}, \check{\theta}) \leq e^{-Cn\underline{\theta}_\star^2 + CJ\sqrt{n\log(p)}}.$$

For $\delta_j^{(k)} = 1$, and $\theta \in \mathsf{B}_{\delta^{(k)}}$, it follows from (24) that

$$\theta_j \langle X_j, y - X\theta_\vartheta\rangle \geq \frac{n\theta_{\star j}^2}{\sigma^2} - C\left(s_\star + \min(J, s_1)\right)\sqrt{n\log(p)} \geq \frac{n\underline{\theta}_\star^2}{2\sigma^2} - CJ\sqrt{n\log(p)},$$

under the sample size condition (19). We deduce that

$$1 - \tilde{q}_j(\vartheta,\theta) \leq \exp\left(\mathsf{a} - \frac{\theta_j}{\sigma^2}\langle X_j, y - X\theta_\vartheta\rangle\right) \leq \exp\left(\mathsf{a} - \frac{n\underline{\theta}_\star^2}{2\sigma^2} + CJ\sqrt{n\log(p)}\right)$$
$$\leq \exp\left(-C_1\sqrt{n}\underline{\theta}_\star + C_2 J\sqrt{\log(p)}\right).$$

The last two majorations on $1 - q_j(\check{\vartheta}^{(r-1)}, \check{\theta})$ and $1 - \tilde{q}_j(\vartheta,\theta)$, and (28) implies that for $\delta_j^{(k)} = \check{\delta}_j^{(k)}$,

$$\int \mathcal{I}_{r,j}^{(k)}(\theta,\check{\theta})\,\Pi(\mathrm{d}\theta|\delta^{(k)},\mathcal{D})\,\Pi(\mathrm{d}\check{\theta}|\check{\delta}^{(k)},\mathcal{D}) \leq e^{-Cn\underline{\theta}_\star^2} + \frac{4}{p^{1-M/2}}. \tag{29}$$

The claim (19) follows from (25) and (29) together.

□

D.1. **Technical lemmas.**

Lemma 1. *Assume H2, and let $\mathcal{E}_0$ as in (14). Let $\{\delta^{(k)},\ k \geq 0\}$ be the δ-marginal chain generated by Algorithm 2 for the linear regression posterior. There exists a constant $C > 0$ that depends only on c (in the definition of $\mathcal{E}_0$) and the constants in H2 such that for*

$$\mathsf{u} \geq C(1 + s_\star^2), \quad \textit{and} \quad n \geq \left(\|\delta^{(0)}\|_0 + s_\star + 2\log(p)\right)^2 \log(p), \tag{30}$$

it holds

$$\mathbf{1}_{\mathcal{E}_0}(Y)\mathbb{P}\left(\|\delta^{(k)}\|_0 > \|\delta^{(0)}\|_0 + s_\star + 2\log(p)\right) \leq \left(\frac{p - s_\star}{2p}\right)^{s - \|\delta^{(0)}\|_0 - s_\star} \leq \frac{1}{p}.$$

Proof. Fix $Y \in \mathcal{E}_0$. Set $s_1 \stackrel{\text{def}}{=} \|\delta^{(0)}\|_0 + s_\star + 2\log(p)$. Referring to the coupling construction at the beginning of the proof, the event $\{\|\delta^{(k)}\|_0 > s_1\}$ means that we can find at least $s_1 - \|\delta^{(0)}\|_0 - s_\star$ terms among $\{(\delta^{(t)}, \mathsf{J}_r^{(t)}, d_r^{(t)}),\ 1 \leq t \leq k-1, 1 \leq r \leq J\}$ where $\|\delta^{(t)}\|_0 \leq s_1$, $\mathsf{J}_r^{(t)} \in \{j :\ \delta_{\star j} = 0, \text{ and } \delta_j^{(t)} = 0\}$, and $d_r^{(t)} = 1$.

$$\mathbb{P}\left(\mathsf{J}_r^{(t)} \in \{j :\ \delta_{\star j} = 0, \text{ and } \delta_j^{(t)} = 0\}|\delta^{(t)}\right) \leq 1 - \frac{s_\star}{p}.$$

We show next that on the event $\|\delta^{(t)}\|_0 \leq s_1$, and $\mathsf{J}_r^{(t)} \in \{j :\ \delta_{\star j} = 0, \text{ and } \delta_j^{(t)} = 0\}$,

$$\mathbb{P}\left(d_r^{(t)} = 1|\delta^{(t)}, \mathsf{J}_r^{(t)} = j\right) \leq \frac{1}{2}, \tag{31}$$

to conclude that

$$\mathbb{P}\left(\|\delta^{(k)}\|_0 > s_1\right) \leq \left(\frac{p - s_\star}{2p}\right)^{s_1 - \|\delta^{(0)}\|_0 - s_\star} \leq \exp\left(-(s_1 - \|\delta^{(0)}\|_0 - s_\star)\log(2)\right) \leq \frac{1}{p},$$

which would end the proof. In order to prove (31), for some absolute constant $m > 0$, let

$$\mathbb{S} \stackrel{\text{def}}{=} \left\{\theta \in \mathbb{R}^p : \ \theta \in \mathsf{B}_{\delta^{(t)}}, \quad \text{and} \quad |\theta_j| \leq \sqrt{\frac{m}{\rho_0}}\right\}.$$

As seen in (17), we can choose m such that $\Pi(\theta \notin \mathbb{S}|\delta^{(t)}, \mathcal{D}) \leq \frac{1}{4}$. Fix j such that $\delta_{\star j} = 0$ and $\delta_j^{(t)} = 0$. Recalling the expression of $\tilde{q}_j$ in (6), it follows then that

$$\mathbb{P}\left(d_r^{(t)} = 1 | \mathsf{J}_r^{(t)} = j, \ \delta^{(t)}\right) \leq \frac{1}{4} +$$
$$\int_{\mathbb{S}} \exp\left(-\mathsf{a} + \frac{\theta_j^2}{2}(\rho_0 - \rho_1) + \frac{\theta_j}{\sigma^2}\langle X_j, y - X\theta_\vartheta\rangle + \frac{\theta_j^2}{2\sigma^4}\langle X_j, y - X\theta_\vartheta\rangle^2\right)\Pi(\theta|\delta^{(t)}, \mathcal{D})\mathrm{d}\theta.$$

Then we write

$$\begin{aligned}\langle X_j, y - X\theta_\vartheta\rangle &= \langle X_j, y - X\theta_\star\rangle + \langle X_j, X\theta_\star - X\theta_\vartheta\rangle \\ &= \langle X_j, y - X\theta_\star\rangle + \sum_{r \neq j}\langle X_j, X_r\rangle(\theta_{\star r} - \theta_r\vartheta_r).\end{aligned}$$

The last summation does not include j because $\theta_{\star j} = 0$, and $\mathsf{J}_r^{(t)} = j$, which implies that $\vartheta_j = 0$. Since $\theta \in \mathbb{S}$, we see that $|\theta_{\star r} - \theta_r\vartheta_r| \leq C$ for all r, for some constant C. If $\delta_{\star r} = 0$, then $|\theta_r| \leq C\sqrt{\log(p)/n}$. It follows that for $Y \in \mathcal{E}_0$,

$$\begin{aligned}|\theta_j\langle X_j, y - X\theta_\vartheta\rangle| &\leq |\theta_j|\left(\sqrt{cn\log(p)} + Cs_\star\sqrt{c_0 n\log(p)} + Cs_1\sqrt{\frac{\log(p)}{n}}\sqrt{c_0 n\log(p)}\right) \\ &\leq C|\theta_j|\left(s_\star\sqrt{n\log(p)} + s_1\log(p)\right) \leq Cs_\star\sqrt{\log(p)},\end{aligned}$$

under the sample size condition $n \geq s_1^2\log(p)$. Hence taking $\mathsf{u} > C(1 + s_\star)^2$ large enough, it follows that

$$\mathbb{P}\left(d_r^{(t)} = 1 | \mathsf{J}_r^{(t)} = j, \ \delta^{(t)}\right) \leq \frac{1}{4} +$$
$$\int_{\mathbb{S}} \exp\left(-\mathsf{u}\log(p) - \frac{1}{2}\log\left(\frac{n}{\sigma^2}\right) + C(1 + s_\star)^2\log(p)\right)\Pi(\theta|\delta^{(t)}, \mathcal{D})\mathrm{d}\theta + \frac{1}{4} \leq \frac{1}{4} + \frac{1}{4} \leq \frac{1}{2}.$$

$\square$

Lemma 2. *Assume H2, and let $\mathcal{E}_0$ as in (14). Fix $0 < s_1 \leq p$. Then we can find constants C, C' that depends only on $\sigma^2, \|\theta_\star\|_\infty, c_0$, and c (in the definition of $\mathcal{E}_0$) such that for $n \geq C s_1^2 \log(p)$, the following holds. For all $\delta \in \Delta$ such that $\|\delta\|_0 \leq s_1$,*

$$\|\hat{\theta}_\delta - [\theta_\star]_\delta\|_\infty \leq C'\left(1 + m(\delta)\right)\sqrt{\frac{\log(p)}{n}}, \tag{32}$$

where $m(\delta) \stackrel{\text{def}}{=} |\{k : \delta_{\star k} = 1, \ \delta_k = 0\}|$, and $\hat{\theta}_\delta$ as in (16).

Proof. The proof follows (Lounici (2008) Theorem 1). Fix $Y \in \mathcal{E}_0$. The first order optimality condition of $\hat{\theta}_\delta$ is given by $-\rho_1\hat{\theta}_\delta + X_\delta'(Y - X_\delta\hat{\theta}_\delta)/\sigma^2 = 0$, which can be rewritten as

$$\left(\rho_1 I_{\|\delta\|_0} + \frac{1}{\sigma^2}X_\delta' X_\delta\right)([\theta_\star]_\delta - \hat{\theta}_\delta) - \rho_1[\theta_\star]_\delta + \frac{1}{\sigma^2}X_\delta'(X\theta_\star - X_\delta[\theta_\star]_\delta) + \frac{1}{\sigma^2}X_\delta'(Y - X\theta_\star) = 0.$$

We deduce that

$$\begin{aligned}\left\|\left(\rho_1 I_{\|\delta\|_0} + \frac{1}{\sigma^2}X_\delta' X_\delta\right)([\theta_\star]_\delta - \hat{\theta}_\delta)\right\|_\infty &\leq \rho_1\|\theta_\star\|_\infty \\ &+ \frac{1}{\sigma^2}\max_{k:\ \delta_k=1}\sum_{j:\ \delta_{\star j}=1, \delta_j=0}|\theta_{\star j}|\left|\langle X_j, X_k\rangle\right| + \frac{1}{\sigma}\sqrt{cn\log(p)} \\ &\leq C(1 + m(\delta))\sqrt{n\log(p)}.\end{aligned}$$

Using this conclusion and the restricted strong convexity in (13), for $n \geq C s_1^2 \log(p)$, we have

$$\begin{aligned}\frac{n}{2}\|\hat{\theta}_\delta - [\theta_\star]_\delta\|_2^2 &\leq (\hat{\theta}_\delta - [\theta_\star]_\delta)'\left(\rho_1 I_{\|\delta\|_0} + \frac{1}{\sigma^2}X_\delta' X_\delta\right)(\hat{\theta}_\delta - [\theta_\star]_\delta) \\ &\leq C(1 + m(\delta))\sqrt{n\log(p)}\|\hat{\theta}_\delta - [\theta_\star]_\delta\|_1 \\ &\leq C s_1^{1/2}(1 + m(\delta))\sqrt{n\log(p)}\|\hat{\theta}_\delta - [\theta_\star]_\delta\|_2,\end{aligned}$$

which implies that

$$\|\hat{\theta}_\delta - [\theta_\star]_\delta\|_2 \leq C(1 + m(\delta))\sqrt{\frac{s_1\log(p)}{n}}.$$

On the other hand for j such that $\delta_j = 1$,

$$\begin{aligned}\left(\left(\rho_1 I_{\|\delta\|_0} + \frac{1}{\sigma^2}X_\delta' X_\delta\right)([\theta_\star]_\delta - \hat{\theta}_\delta)\right)_j &= (\rho_1 + \frac{n}{\sigma^2})\left(\hat{\theta}_\delta - [\theta_\star]_\delta\right)_j \\ &+ \frac{1}{\sigma^2}\sum_{k\neq j:\ \delta_k=1}\langle X_k, X_j\rangle\left(\hat{\theta}_\delta - [\theta_\star]_\delta\right)_k,\end{aligned}$$

which we use to deduce that

$$\|\hat\theta_\delta-[\theta_\star]_\delta\|_\infty \leq \frac{\sigma^2}{n}\left\|\left(\rho_1 I_{\|\delta\|_0}+\frac{1}{\sigma^2}X_\delta' X_\delta\right)([\theta_\star]_\delta-\hat\theta_\delta)\right\|_\infty + \frac{1}{n}\sqrt{c_0 n\log(p)}\|\hat\theta_\delta-[\theta_\star]_\delta\|_1$$

$$\leq C(1+m(\delta))\sqrt{\frac{\log(p)}{n}} + s_1^{1/2}\sqrt{\frac{c_0\log(p)}{n}}\|\hat\theta_\delta-[\theta_\star]_\delta\|_2$$

$$\leq C(1+m(\delta))\sqrt{\frac{\log(p)}{n}} + C(1+m(\delta))\sqrt{\frac{\log(p)}{n}}\sqrt{\frac{s_1^2\log(p)}{n}}$$

$$\leq C(1+m(\delta))\sqrt{\frac{\log(p)}{n}},$$

under the stated sample size condition. □

We show in the next result that the posterior distribution puts most of its probability mass on models that contain the true model $\delta_\star$.

Lemma 3. *Assume H2, and let $\mathcal{E}_{s_\star}$ be as in (14). Then we can find constants C_1, C_2 that depends only on $\sigma^2, \|\theta_\star\|_\infty, c_0$, c_1 c_2 and c (in the definition of $\mathcal{E}_{s_\star}$) such that for $n \geq C_1\underline{\theta}_\star^{-2}(1+s_\star^3)\log(p)$, it holds*

$$\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\Pi(\mathcal{C}|\mathcal{D})\right] \geq 1-\frac{3}{p},$$

where

$$\mathcal{C} \stackrel{\text{def}}{=} \{\delta\in\Delta:\ \delta\supseteq\delta_\star,\ \ \textit{and}\ \ \|\delta\|_0\leq C_2(1+s_\star)\}.$$

Proof. By Lemma 5, there exist positive constant C_1, C_2 that depends only on c_0, c_2, and c such that for $n \geq C_1 s_\star^2\log(p)$,

$$\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_0}(Y)\Pi\left(\|\delta\|_0 > C_2(1+s_\star)\mid\mathcal{D}\right)\right] \leq \frac{2}{p}.$$

We set

$$s \stackrel{\text{def}}{=} C_2(1+s_\star),$$

and $\mathcal{A} \stackrel{\text{def}}{=} \{\delta:\ \delta\not\supseteq\delta_\star,\ \|\delta\|_0\leq s\}$, so that

$$\Delta = \mathcal{C}_s\cup\mathcal{A}\cup\{\delta\in\Delta:\ \|\delta\|_0 > s\}.$$

Therefore,

$$\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\Pi(\mathcal{C}_s|\mathcal{D})\right] \geq 1-\frac{2}{p}-\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_{s_\star}}(Y)\Pi\left(\mathcal{A}|\mathcal{D}\right)\right].$$

To finish the proof we will establish that for $Y\in\mathcal{E}_{s_\star}$, $\Pi\left(\mathcal{A}|\mathcal{D}\right)\leq\frac{1}{p}$. To that end, let $\mathcal{P}\stackrel{\text{def}}{=}\{\delta^{(0)}\in\Delta:\ \delta^{(0)}\subseteq\delta_\star,\ \delta^{(0)}\neq\delta_\star\}$, and for each $\delta^{(0)}\in\mathcal{P}$, we set

$$\mathcal{A}(\delta^{(0)}) \stackrel{\text{def}}{=} \left\{\delta\in\Delta:\ \delta\supseteq\delta^{(0)},\ \|\delta\|_0\leq s,\ \text{and}\ \delta_{\star j}=0\ \text{whenever}\ (\delta_j=1,\ \text{and}\ \delta_j^{(0)}=0)\right\}.$$

We then write

$$\Pi(\mathcal{A}|\mathcal{D}) = \Pi(\delta_\star|\mathcal{D}) \sum_{\delta^{(0)}\in\mathcal{P}} \frac{\Pi(\delta^{(0)}|\mathcal{D})}{\Pi(\delta_\star|\mathcal{D})} \sum_{\delta\in\mathcal{A}(\delta^{(0)})} \frac{\Pi(\delta|\mathcal{D})}{\Pi(\delta^{(0)}|\mathcal{D})}. \tag{33}$$

For any subset $\delta, \vartheta \in \Delta$, we have

$$\frac{\Pi(\delta|\mathcal{D})}{\Pi(\vartheta|\mathcal{D})} = \left(\frac{1}{p^{\mathsf{u}}}\sqrt{\frac{\rho_1}{2\pi}}\right)^{\|\delta\|_0-\|\vartheta\|_0} \frac{\int_{\mathbb{R}^{\|\delta\|_0}} \exp\left(-\frac{\rho_1}{2}\|u\|_2^2 - \frac{1}{2\sigma^2}\|Y - X_\delta u\|_2^2\right) \mathrm{d}u}{\int_{\mathbb{R}^{\|\vartheta\|_0}} \exp\left(-\frac{\rho_1}{2}\|u\|_2^2 - \frac{1}{2\sigma^2}\|Y - X_\vartheta u\|_2^2\right) \mathrm{d}u}.$$

We calculate that for any $\delta \in \Delta$,

$$\int_{\mathbb{R}^{\|\delta\|_0}} \exp\left(-\frac{\rho_1}{2}\|u\|_2^2 - \frac{1}{2\sigma^2}\|Y - X_\delta u\|_2^2\right) \mathrm{d}u = \left(\frac{2\pi}{\rho_1}\right)^{\|\delta\|_0/2} \frac{e^{-\frac{1}{2\sigma^2}Y'\left(I_n + \frac{1}{\rho_1\sigma^2}X_\delta X_\delta'\right)^{-1}Y}}{\sqrt{\det\left(I_n + \frac{1}{\rho_1\sigma^2}X_\delta X_\delta'\right)}}.$$

And we deduce, using $\rho_1 = 1$, that

$$\frac{\Pi(\delta|\mathcal{D})}{\Pi(\vartheta|\mathcal{D})} = \left(\frac{1}{p^{\mathsf{u}}}\right)^{\|\delta\|_0-\|\vartheta\|_0} e^{\frac{1}{2\sigma^2}\left(Y'L_\vartheta^{-1}Y - Y'L_\delta^{-1}Y\right)} \sqrt{\frac{\det(L_\vartheta)}{\det(L_\delta)}}, \tag{34}$$

where

$$L_\delta \stackrel{\text{def}}{=} I_n + \frac{1}{\sigma^2}X_\delta X_\delta'.$$

Suppose that $\vartheta \supseteq \delta$, and $\|\vartheta\|_0 \le s_1$. In that case

$$L_\vartheta = L_\delta + \frac{1}{\sigma^2}X_{\vartheta-\delta}X_{\vartheta-\delta}'.$$

and by the determinant lemma ($\det(A + UV') = \det(A)\det(I_m + V'A^{-1}U)$ valid for any invertible matrix $A \in \mathbb{R}^{n\times n}$, and $U, V \in \mathbb{R}^{n\times m}$), and using the lower bound on the smallest eigenvalue of L_δ resulting from (13), we have

$$1 \le \frac{\det(L_\vartheta)}{\det(L_\delta)} = \det\left(I_{\vartheta-\delta} + \frac{1}{\sigma^2}X_{\vartheta-\delta}'L_\delta^{-1}X_{\vartheta-\delta}\right) \le (1 + 2\|\vartheta-\delta\|_0)^{\|\vartheta-\delta\|_0}.$$

We use this to deduce from (34) that when $\vartheta \supseteq \delta$, and $\|\vartheta\|_0 \le s$, it holds

$$\begin{aligned} p^{\mathsf{u}\|\vartheta-\delta\|_0} e^{\frac{1}{2\sigma^2}\left(Y'L_\vartheta^{-1}Y - Y'L_\delta^{-1}Y\right)} &\le \frac{\Pi(\delta|\mathcal{D})}{\Pi(\vartheta|\mathcal{D})} \\ &\le p^{\mathsf{u}\|\vartheta-\delta\|_0}(1+2s)^{\|\vartheta-\delta\|_0} e^{\frac{1}{2\sigma^2}\left(Y'L_\vartheta^{-1}Y - Y'L_\delta^{-1}Y\right)}. \end{aligned} \tag{35}$$

By the Woodbury formula which states that any set of matrices U, V, A, C with matching dimensions, $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, we have

$$Y'L_\vartheta^{-1}Y - Y'L_\delta^{-1}Y = -\frac{1}{\sigma^2}y'L_\delta^{-1}X_{\vartheta-\delta}\left(I_{\|\vartheta-\delta\|_0} + \frac{1}{\sigma^2}X_{\vartheta-\delta}'L_\delta^{-1}X_{\vartheta-\delta}\right)^{-1} X_{\vartheta-\delta}L_\delta^{-1}y.$$

It follows from Equation (37) of Lemma 4 that for any non-zero vector $u \in \mathbb{R}^{\|\vartheta-\delta\|_0}$,

$$u'X'_{\vartheta-\delta}L_\delta^{-1}X_{\vartheta-\delta}u \geq \frac{n}{2}\|u\|_2^2 - C_0\sqrt{n\log(p)}\|u\|_1^2 \geq \frac{n}{4}\|u\|_2^2,$$

for some absolute constant C_0, provided that $n \geq 4C_0^2 s^2 \log(p)$. We deduce that for $\delta \subseteq \vartheta$, $\|\vartheta\|_0 \leq s$, it holds

$$\frac{\|Y'L_\delta^{-1}X_{\vartheta-\delta}\|_2^2}{\sigma^2(1+\|\vartheta-\delta\|_0 n)} \leq Y'L_\delta^{-1}Y - Y'L_\vartheta^{-1}Y \leq \frac{4\|Y'L_\delta^{-1}X_{\vartheta-\delta}\|_2^2}{\sigma^2 n}. \tag{36}$$

We put (36) and (35) to write the second summation of (33) as

$$\sum_{\delta\in\mathcal{A}(\delta^{(0)})}\frac{\Pi(\delta|\mathcal{D})}{\Pi(\delta^{(0)}|\mathcal{D})} = \sum_{k=0}^{s-\|\delta^{(0)}\|_0}\sum_{\delta\in\mathcal{A}(\delta^{(0)}):\ \|\delta\|_0=\|\delta^{(0)}\|_0+k}\frac{\Pi(\delta|\mathcal{D})}{\Pi(\delta^{(0)}|\mathcal{D})}$$
$$\leq \sum_{k=0}^{s-\|\delta^{(0)}\|_0}\sum_{\delta\in\mathcal{A}(\delta^{(0)}):\ \|\delta\|_0=\|\delta^{(0)}\|_0+k}\left(\frac{1}{p^{\mathsf{u}}}\right)^k\exp\left(\frac{2\left\|Y'L_{\delta^{(0)}}^{-1}X_{\delta-\delta^{(0)}}\right\|_2^2}{\sigma^4 n}\right).$$

We can write $Y = \sigma V + \sum_{k:\ \delta_{\star k}=1}\theta_{\star k}X_k$, where $V = (Y - X\theta_\star)/\sigma$. Fix a component i such that $(\delta-\delta^{(0)})_i = 1$. Note that we have $\delta_i^{(0)} = 0$, and $\delta_{\star i} = 0$. We can then write

$$Y'L_{\delta^{(0)}}^{-1}X_i = \sigma V'L_{\delta^{(0)}}^{-1}X_i + \sum_{k:\ \delta_{\star k}=1}\theta_{\star k}X'_kL_{\delta^{(0)}}^{-1}X_i.$$

Therefore, by (37) from Lemma 4, we have

$$|Y'L_{\delta^{(0)}}^{-1}X_i| \leq C_1\sqrt{(1+s_\star)n\log(p)} + C_1(s_\star - \|\delta^{(0)}\|_0)\sqrt{n\log(p)}.$$

It follows that

$$\frac{\left\|Y'L_{\delta^{(0)}}^{-1}X_{\delta-\delta^{(0)}}\right\|_2^2}{\sigma^4 n} \leq C_1 k(1+s_\star^2)\log(p),$$

for some constant C_1. Therefore,

$$\sum_{\delta\in\mathcal{A}(\delta^{(0)})}\frac{\Pi(\delta|\mathcal{D})}{\Pi(\delta^{(0)}|\mathcal{D})} \leq p^{C_1 s(1+s_\star^2)}\sum_{k=0}^{s-\|\delta^{(0)}\|_0}\binom{p-\|\delta^{(0)}\|_0}{k}\left(\frac{1}{p^{\mathsf{u}}}\right)^k \leq 2p^{C_1 s(1+s_\star^2)},$$

by choosing $\mathsf{u} > 2$, assuming $p \geq 2$. The last display, with (33) and (35) yield

$$\Pi(\mathcal{A}|\mathcal{D}) \leq 2p^{C_1 s(1+s_\star^2)}\sum_{\delta^{(0)}\in\mathcal{P}}\frac{\Pi(\delta^{(0)}|\mathcal{D})}{\Pi(\delta_\star|\mathcal{D})}$$
$$\leq 2p^{C_1 s(1+s_\star^2)}\sum_{k=0}^{s_\star-1}\sum_{\delta^{(0)}\in\mathcal{P}:\ \|\delta^{(0)}\|_0=s_\star-k}p^{k\mathsf{u}}(1+2s)^k\exp\left(-\frac{\left\|Y'L_{\delta^{(0)}}^{-1}X_{\delta_\star-\delta^{(0)}}\right\|_2^2}{2\sigma^4(1+kn)}\right).$$

As above, given i such that $\delta_{\star i} = 1$, we write

$$Y' L_{\delta^{(0)}}^{-1} X_i = \sigma V' L_{\delta^{(0)}}^{-1} X_i + \theta_{\star i} X_i' L_{\delta^{(0)}}^{-1} X_i + \sum_{k \neq i:\ \delta_{\star k}=1} \theta_{\star k} X_k' L_{\delta^{(0)}}^{-1} X_i,$$

and using Lemma 4, we deduce that

$$|Y' L_{\delta^{(0)}}^{-1} X_i| \geq \frac{n|\theta_{\star i}|}{2} - C_1 \sqrt{(1+s_\star) n \log(p)} \geq \frac{n|\underline{\theta}_\star|}{4},$$

under the sample size condition $n \geq C_2(1+s_\star)\log(p)/\underline{\theta}_\star^2$. Hence

$$\frac{\left\| Y' L_{\delta^{(0)}}^{-1} X_{\delta_\star - \delta^{(0)}} \right\|_2^2}{2\sigma^4(1+kn)} \geq \frac{n\underline{\theta}_\star^2}{64\sigma^4},$$

so that

$$\begin{aligned}
\Pi(\mathcal{A}|\mathcal{D}) &\leq 2p^{C_1 s(1+s_\star^2)} e^{-\frac{n\underline{\theta}_\star^2}{64\sigma^4}} \sum_{k=0}^{s_\star - 1} \binom{s_\star}{k} (1+2s)^k \\
&\leq 2p^{C_1 s(1+s_\star^2)} e^{-\frac{n\underline{\theta}_\star^2}{64\sigma^4}} \sum_{k=0}^{s_\star - 1} (C_1(1+s_\star))^k \\
&\leq 2p^{C_1 s(1+s_\star^2)} e^{-\frac{n\underline{\theta}_\star^2}{64\sigma^4}} (C_1(1+s_\star))^{s_\star} \\
&\leq 2\exp\left(-\frac{n\underline{\theta}_\star^2}{64\sigma^4} + C_1(1+s_\star^3)\log(p)\right) \\
&\leq \exp\left(-\frac{n\underline{\theta}_\star^2}{128\sigma^4}\right) \leq \frac{1}{p},
\end{aligned}$$

for $n \geq C_2 \underline{\theta}_\star^{-2}(1+s_\star^3)\log(p)$, for some constant C_2. This ends the proof.

□

Lemma 4. *Assume H2, and fix $s > 0$. There exist constants C_1, C_2 that depends only on $\sigma, \underline{\kappa}, c_0$ and $\|\theta_\star\|_\infty$ such that for $n \geq C_1 s^2 \log(p)$, the following holds. For all $\delta \in \Delta_s$, and for all pair $j \neq k$, such that $\delta_j = 0$, it holds*

$$|X_j' L_\delta^{-1} X_k| \leq C_2 \left(1 + \frac{\|\delta\|_0^{1/2}}{n} \mathbf{1}_{\{\delta_k = 1\}}\right) \sqrt{n\log(p)}, \quad \text{and} \quad X_j' L_\delta^{-1} X_j \geq \frac{n}{2}. \tag{37}$$

Proof. Applying the Woodbury identity to L_δ, we have

$$L_\delta^{-1} = I_n - \frac{1}{\sigma^2} X_\delta \left(I_{\|\delta\|_0} + \frac{1}{\sigma^2} X_\delta' X_\delta \right)^{-1} X_\delta'. \tag{38}$$

It follows that

$$X_j' L_\delta^{-1} X_k = \langle X_j, X_k \rangle - \frac{1}{\sigma^2} X_j' X_\delta \left(I_{\|\delta\|_0} + \frac{1}{\sigma^2} X_\delta' X_\delta \right)^{-1} X_\delta' X_k.$$

Under the sample size condition, by (13), we have

$$\left|\frac{1}{\sigma^2}X_j'X_\delta\left(I_{\|\delta\|_0}+\frac{1}{\sigma^2}X_\delta'X_\delta\right)^{-1}X_\delta'X_k\right| \leq \frac{2\|X_\delta'X_j\|_2\|X_\delta'X_k\|_2}{\sigma^2 n}.$$

By assumption $\delta_j = 0$. If $\delta_k = 0$, then

$$\|X_\delta'X_j\|_2\|X_\delta'X_k\|_2 \leq \sqrt{c_0\|\delta\|_0 n\log(p)}\sqrt{c_0\|\delta\|_0 n\log(p)},$$

and we deduce that

$$|X_j'L_\delta^{-1}X_k| \leq \sqrt{c_0 n\log(p)} + \frac{2c_0\|\delta\|_0\log(p)}{\sigma^2} \leq C\sqrt{n\log(p)},$$

provided that $n \geq s^2\log(p)$. Suppose now that $\delta_k = 1$. Note that starting from (38) we can also write

$$X_\delta'L_\delta^{-1}X_j = \left(I_{\|\delta\|_0}+\frac{1}{\sigma^2}X_\delta'X_\delta\right)^{-1}X_\delta'X_j.$$

This implies that if $\delta_k = 1$, then

$$|X_j'L_\delta^{-1}X_k| \leq \frac{2\|X_\delta'X_j\|_2}{n} \leq \frac{2\sqrt{c_0\|\delta\|_0 n\log(p)}}{n} \leq C_2\sqrt{\frac{\|\delta\|_0\log(p)}{n}},$$

which establishes the first part of (37). When $j = k$, we get

$$X_j'L_\delta^{-1}X_j \geq n - \frac{2c_0\|\delta\|_0\log(p)}{\sigma^2} \geq \frac{n}{2},$$

under the sample size condition $n \geq 2c_0 s\log(p)/\sigma^2$.

□

Lemma 5. *Assume H2, and let $\mathcal{E}_0$ as in (14). There exist positive constant C_1, C_2 that depends only on c_0, c_2, and c (in the definition of $\mathcal{E}_0$) such that for $n \geq C_1 s_\star^2\log(p)$,*

$$\mathbb{E}_\star\left[\mathbf{1}_{\mathcal{E}_0}(Y)\Pi\left(\|\delta\|_0 > (1+C_2)s_\star \mid \mathcal{D}\right)\right] \leq \frac{2}{p}.$$

Proof. The lemma follows from Theorem 2.2 of Atchade and Bhattacharyya (2019), applied with $\bar{\rho} = 2\sqrt{cn\log(p)}/\sigma$, and $\bar{\kappa} = s_\star n$. The sub-Gaussian assumption in H2-(1) implies that Equation (2.1) of Atchade and Bhattacharyya (2019) holds with $\mathsf{r}_0 = n/(2\sigma^2\bar{\rho})$ under the sample size condition. Then using the assumption in H2 that n/p and $\|\theta_\star\|_\infty/\log(p)$ remain bounded from above by c_2, we checked that Equation 2.2 of Atchade and Bhattacharyya (2019) is satisfies for some absolute constant c_0. □

Appendix E. Description of the coupled chains for mixing time estimation

We describe here the specific coupled Markov chain employed to estimate the mixing time plots presented in Section 4.1. We describe the method for Algorithm 1. Algorithm 2 proceeds similarly.

We start with a brief description of the method. Let $\{X^{(t)},\ t \geq 0\}$ be the Markov chain generated by Algorithm 1, where $X^{(t)} = (\delta^{(t)}, \theta^{(t)}) \in \mathsf{X}$. Let K denote the transition kernel of the Markov chain $\{X^{(t)},\ t \geq 0\}$. The basic idea of the method is to construct a coupling $\check{K}$ of K with itself: that is, a transition kernel on $\mathsf{X}\times\mathsf{X}$ such that $\check{K}((x,y), A\times\mathsf{X}) = K(x,A)$, $\check{K}((x,y),\mathsf{X}\times B) = K(y,B)$, for all $x, y \in \mathsf{X}$, and all measurable sets A, B. The coupling $\check{K}$ is constructed in such a way that $\check{K}((x,x),\mathcal{D}) = 1$, where $\mathcal{D} \stackrel{\text{def}}{=} \{(x,x):\ x \in \mathsf{X}\}$. The method then proceeds as follows. Fix a lag $L \geq 1$. Draw $X^{(0)} \sim \Pi^{(0)}$, $Y^{(0)} \sim \Pi^{(0)}$ (where $\Pi^{(0)}$ is the initial distribution as given in the initialization step in Algorithm 1). Draw $X^{(L)}|(X^{(0)}, Y^{(0)}) \sim K^L(X^{(0)}, \cdot)$. Then for any $k \geq 1$, draw,

$$(X^{(L+k)}, Y^{(k)})| \left\{(X^{(L+k-1)}, Y^{(k-1)}), \ldots, (X^{(L)}, Y^{(0)})\right\} \sim \check{K}\left((X^{(L+k-1)}, Y^{(k-1)}), \cdot\right).$$

Setting

$$\tau^{(L)} \stackrel{\text{def}}{=} \inf\left\{k > L:\ X^{(k)} = Y^{(k-L)}\right\},$$

it then holds under some ergodicity assumptions on P (see Biswas et al. (2019)) that

$$\|\Pi^{(t)} - \Pi\|_{\mathsf{tv}} \leq \mathbb{E}\left[\max\left(0, \left\lceil \frac{\tau^{(L)} - L - t}{L}\right\rceil\right)\right], \tag{39}$$

where $\lceil x\rceil$ denote the smallest integer above x. The implication of (39) is that we can empirically upper bound the left hand side of (39) by simulating multiple copies of the joint chain as described above and then approximating the expectation on the right hand side of (39) by Monte Carlo. We refer the reader to Biswas et al. (2019) for more details on the construction of such coupled kernels.

We modify Algorithm 1 to construct the coupled kernel $\check{P}$. Let $(\delta^{(1,t)}, \theta^{(1,t)})$ and let $(\delta^{(2,t)}, \theta^{(2,t)})$ denote the states of the two chains at time t. At some iteration $t \geq 1$, given $(\delta^{(1,L+t)}, \theta^{(1,L+t)}) = (\delta^{(1)}, \theta^{(1)})$ and $(\delta^{(2,t)}, \theta^{(2,t)}) = (\delta^{(2)}, \theta^{(2)})$, we now describe how to generate the next state of the coupled chain.

In step 1, to update $\delta^{(1)}$ and $\delta^{(2)}$, we first make use of the same randomly drawn subset J. For $i = 1, 2$, drawing $\bar{\delta}^{(i)} \sim Q_\theta^{(\mathsf{J})}(\delta^{(i)}, \cdot)$ is equivalent to let $\bar{\delta}^{(i)}_{-\mathsf{J}} = \delta^{(i)}_{-\mathsf{J}}$, and for any $j \in \mathsf{J}$, draw $\bar{\delta}^{(i)}_j \sim \mathbf{Ber}(q_j^{(i)})$ which we implement in the following way. We first draw a common uniform number $u_j \sim \mathbf{Uniform}(0,1)$, then we obtain $\bar{\delta}^{(i)}_j = \mathbf{1}\{q_j^{(i)} \leq u_j\}$ for $i = 1, 2$.

In step 2, to update $\theta^{(1)}$ and $\theta^{(2)}$, for simplicity, we partition the indices $\{1,\ldots,p\}$ into four groups: $G_{ab} = \{j : \bar{\delta}_j^{(1)} = a, \bar{\delta}_j^{(2)} = b\}$ for $a, b = 0, 1$.

To update the components of $\theta^{(1)}_{G_{00}}$ and $\theta^{(2)}_{G_{00}}$, for any $j \in G_{00}$ we first draw a common standard normal random variables Z_j, and then obtain $\bar{\theta}_j^{(i)} = \rho_0^{-\frac{1}{2}} Z_j$ for $i = 1, 2$.

To update the components of $\theta^{(1)}_{G_{01}}$ and $\theta^{(2)}_{G_{01}}$, Since in linear regression, $[\theta]_\delta \mid \delta \sim \mathbf{N}(\hat{\theta}_\delta, \Sigma)$, where $\hat{\theta}_\delta$ is describled in (16) and $\Sigma = \sigma^2(\sigma^2\rho_1 I_{\|\delta\|_0})^{-1}$, we then have $\theta^{(1)} \mid \delta^{(1)} \sim \mathbf{N}(\hat{\theta}^{(1)}, \Sigma^{(1)})$ and $\theta^{(2)} \mid \delta^{(2)} \sim \mathbf{N}(\hat{\theta}^{(2)}, \Sigma^{(2)})$, respectively. Then with the property of gaussian random variables, we have $\theta^{(2)}_{G_{01}} \sim \mathbf{N}(\hat{\theta}^{(2)}_{G_{01}}, \Sigma^{(2)}_{G_{01}})$, where $\hat{\theta}^{(2)}_{G_{01}}$ are the G_{01} components of $\hat{\theta}^{(2)}$ and $\Sigma^{(2)}_{G_{01}}$ is the submatrix of $\Sigma^{(2)}$ with G_{01} rows and columns. With $\theta^{(1)}_{G_{01}} \sim \mathbf{N}(\mathbf{0}, \rho_0^{-1} I_{\|\delta_{G_{01}}\|_0})$, we draw the maximal coupling of these two gaussian distributions to update $\bar{\theta}^{(1)}_{G_{01}}$ and $\bar{\theta}^{(2)}_{G_{01}}$. A similar updating procedure is used for the components of $\bar{\theta}^{(1)}_{G_{10}}$ and $\bar{\theta}^{(2)}_{G_{10}}$.

For components of $\theta^{(1)}_{G_{11}}$ and $\theta^{(2)}_{G_{11}}$, since we have $\theta^{(1)}_{G_{11}} \sim \mathbf{N}(\hat{\theta}^{(1)}_{G_{11}}, \Sigma^{(1)}_{G_{11}})$, where $\hat{\theta}^{(1)}_{G_{11}}$ are the G_{11} components of $\hat{\theta}^{(1)}$ and $\Sigma^{(1)}_{G_{11}}$ is the submatrix of $\Sigma^{(1)}$ with G_{11} rows and columns, and similarly $\theta^{(2)}_{G_{11}} \sim \mathbf{N}(\hat{\theta}^{(2)}_{G_{11}}, \Sigma^{(2)}_{G_{11}})$, we could construct another maximal coupling to update $\bar{\theta}^{(1)}_{G_{11}}$ and $\bar{\theta}^{(2)}_{G_{11}}$.

References

Atchade, Y. and Bhattacharyya, A. (2019). An approach to large-scale quasi-bayesian inference with spike-and-slab priors.

Biswas, N., Jacob, P. E. and Vanetti, P. (2019). Estimating convergence of markov chains with l-lag couplings.

De Sa, C., Olukotun, K. and Ré, C. (2016). Ensuring rapid mixing and low bias for asynchronous gibbs sampling. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*. ICML'16.

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electronic Journal of Statistics* **2** 90 – 102.

Meyn, S. and Tweedie, R. L. (2009). *Markov chains and stochastic stability.* 2nd ed. Cambridge University Press, Cambridge.