SUPPLEMENTARY MATERIALS: Approximate Spectral Gaps for Markov Chains Mixing Times in High Dimensions*

Yves Atchade†

SM1. Extension to reversible Markov semigroups. The idea can also be applied to continuous-time Markov processes. We refer the reader to ([SM2]) for an introduction to Markov semigroups. We consider a reversible Markov semigroup $K = \{K_t,\ t \geq 0\}$, where for each t, K_t is a Markov kernel on $(\mathcal{X}, \mathcal{B})$ that is reversible with respect to π. Let G denote the generator of the semi-group that we assumed well-defined on a dense subspace $\mathcal{A}$ of $L^2(\pi)$ that is stable under G and K_t such that for all $t \geq 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t} K_t f = K_t G f = G K_t f, \quad f \in \mathcal{A}. \tag{SM1.1}$$

We make also the assumption that the domain $\mathcal{A}$ contains constant functions The Dirichlet form of K is defined as

$$\mathcal{E}(f, f) \stackrel{\text{def}}{=} -\int_{\mathcal{X}} f(x) G f(x) \pi(\mathrm{d}x).$$

For $\zeta \in [0, 1)$, we can define the ζ-spectral gap of the semi-group K as

$$\lambda_\zeta(K) \stackrel{\text{def}}{=} \inf \left\{ \frac{-\int_{\mathcal{X}} f(x) G f(x) \pi(\mathrm{d}x)}{\mathsf{Var}_\pi(f) - \zeta},\ \ f \in \mathcal{A},\ \mathsf{Var}_\pi(f) > \zeta,\ \text{and}\ \|f\|_\infty = 1 \right\}. \tag{SM1.2}$$

We have the analog of Lemma 2.1.

Lemma SM1.1. *Let $\nu(\mathrm{d}x) = f(x)\pi(\mathrm{d}x)$ be a probability measure on $\mathcal{X}$, where $f \in \mathcal{A}$. Let $\zeta \in [0, 1)$ be such that $\lambda_\zeta(K) > 0$. Then for all $t \geq 0$ we have*

$$\|\nu K_t - \pi\|_{\mathrm{tv}}^2 \leq \mathsf{Var}_\pi(K_t f) \leq \mathsf{Var}_\pi(f) e^{-2\lambda_\zeta(K)t} + \zeta \|f\|_\infty^2.$$

Remark SM1.2. This result can be applied to Langevin diffusion processes. Suppose that $\mathcal{X} = \mathbb{R}^p$ equipped with the Lebesgue measure, and $\pi(\mathrm{d}x) = e^{-U(x)}/Z$, for a function $U : \mathbb{R}^p \to \mathbb{R}$ that is differentiable with Lipschitz gradient. The Langevin diffusion process for π defines a reversible Markov semigroup with invariant distribution π. The convergence rate of the semigroup toward π is a key ingredient in the analysis of several recent MCMC algorithms, including the unadjusted Langevin algorithm and stochastic gradient Langevin dynamics ([SM7, SM5]). When U is convex, the semigroup is known to possess a spectral gap ([SM3]). Various extensions beyond the convex case are known but typically requires drift conditions ([SM1]). Lemma SM1.1 offers another route, one that might be more effective when a good initial distribution is available, and π has well-understood concentration properties. We leave the details as possible future research.

*Supplementary material for SIMODS #M128308.
https://doi.org/10.1137/19M1283082

†Department of Mathematics and Statistics, Boston University, Boston, MA 02215 (atchade@bu.edu).

Proof. Take $f \in \mathcal{A}$. Without any loss of generality we assume that $\pi(f) = 0$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Var}_\pi(K_t f) = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{X}} (K_t f)^2(x)\pi(\mathrm{d}x) = 2\int_{\mathcal{X}} K_t f(x) G K_t f(x)\pi(\mathrm{d}x). \tag{SM1.3}$$

Suppose that $\|K_t f\|_\infty > 0$. If $\mathsf{Var}_\pi(K_t f/\|K_t f\|_\infty) > \zeta$, then from (SM1.3) and the definition of $\lambda_\zeta(K)$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Var}_\pi(K_t f) &\leq -2\|K_t f\|_\infty^2 \lambda_\zeta(K)\left(\mathsf{Var}_\pi\left(\frac{K_t f}{\|K_t f\|_\infty}\right) - \zeta\right) \\ &\leq -2\lambda_\zeta(K)\mathsf{Var}_\pi(K_t f) + 2\zeta\lambda_\zeta(K)\|K_t f\|_\infty^2.\end{aligned} \tag{SM1.4}$$

However, if $\mathsf{Var}_\pi(K_t f/\|K_t f\|_\infty) \leq \zeta$, we see that the right-hand side of (SM1.4) is nonnegative, whereas from (SM1.3) and the properties of the generator we see that the left-hand side of (SM1.4) is nonpositive. Note also that (SM1.4) continue to hold when $\|K_t f\|_\infty = 0$. Hence for all $f \in \mathcal{A}$, and for all $t \geq 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Var}_\pi(K_t f) \leq -2\lambda_\zeta(K)\mathsf{Var}_\pi(K_t f) + 2\zeta\lambda_\zeta(K)\|f\|_\infty^2. \tag{SM1.5}$$

The lemma then follows from Gronwall's lemma. More precisely, set $\alpha = \zeta\|f\|_\infty^2$, $\beta = 2\lambda_\zeta(K)$, and $u(t) = \mathsf{Var}_\pi(K_t f)$. Hence (SM1.5) reads $u'(t) \leq -\beta u(t) + \alpha\beta$. Setting $v(t) = e^{-\beta t}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{u(t)}{v(t)}\right) = \frac{u'(t)v(t) - v'(t)u(t)}{v(t)^2} = \frac{u'(t) + \beta u(t)}{v(t)} \leq \alpha\beta e^{\beta t}.$$

Integrating both sides yields the stated bound. ■

SM2. Proof of Theorem 4.3.

We start with some basic calculations on the model.

Lemma SM2.1. *Assume H2. For $\delta, \vartheta \in \Delta$ such that $\vartheta \supseteq \delta$, setting $\tau \stackrel{\text{def}}{=} \frac{1}{\sigma^2}\left(\frac{1}{\rho_1} - \frac{1}{\rho_0}\right)$, we have*

$$\frac{\Pi(\vartheta|z)}{\Pi(\delta|z)} = \left(\frac{1}{p^u}\right)^{\|\vartheta\|_0 - \|\delta\|_0} \frac{e^{\frac{\tau}{2\sigma^2} z' L_\delta^{-1} X_{(\vartheta-\delta)}\left(I_{\|\vartheta-\delta\|_0} + \tau X_{(\vartheta-\delta)}' L_\delta^{-1} X_{(\vartheta-\delta)}\right)^{-1} X_{(\vartheta-\delta)}' L_\delta^{-1} z}}{\sqrt{\det\left(I_{\|\vartheta-\delta\|_0} + \tau X_{(\vartheta-\delta)}' L_\delta^{-1} X_{(\vartheta-\delta)}\right)}}. \tag{SM2.1}$$

Proof. For any $\delta \in \Delta_s$,

$$\Pi(\delta|z) \propto \left(\frac{\mathsf{q}}{1-\mathsf{q}}\right)^{\|\delta\|_0} \int_{\mathbb{R}^p} \frac{e^{-\frac{1}{2}\theta' D_{(\delta)}^{-1}\theta}}{\sqrt{\det\left(2\pi D_{(\delta)}\right)}} e^{-\frac{1}{2\sigma^2}\|z - X\theta\|_2^2} \mathrm{d}\theta.$$

where $D_{(\delta)} \in \mathbb{R}^{p\times p}$ is a diagonal matrix with j-th diagonal element equal to ρ_1^{-1} if $\delta_j = 1$, and

ρ_0^{-1} if $\delta_j = 0$. Since $\mathsf{q}/(1-\mathsf{q}) = p^{-u}$, simple algebra gives for any $\vartheta, \delta \in \Delta_s$

$$\frac{\Pi(\vartheta|z)}{\Pi(\delta|z)} = \left(\frac{1}{p^u}\right)^{\|\vartheta\|_0 - \|\delta\|_0} \left(\frac{\rho_1}{\rho_0}\right)^{\frac{\|\vartheta\|_0 - \|\delta\|_0}{2}} \frac{\int_{\mathbb{R}^p} e^{-\frac{1}{2\sigma^2}\|z - Xu\|_2^2 - \frac{1}{2}u' D_{(\vartheta)}^{-1} u} \mathrm{d}u}{\int_{\mathbb{R}^p} e^{-\frac{1}{2\sigma^2}\|z - Xu\|_2^2 - \frac{1}{2}u' D_{(\delta)}^{-1} u} \mathrm{d}u}.$$

$$= \left(\frac{1}{p^u}\right)^{\|\vartheta\|_0 - \|\delta\|_0} \left(\frac{\rho_1}{\rho_0}\right)^{\frac{\|\vartheta\|_0 - \|\delta\|_0}{2}} \frac{\sqrt{\det\left(\sigma^2 D_{(\delta)}^{-1} + X'X\right)}}{\sqrt{\det\left(\sigma^2 D_{(\vartheta)}^{-1} + X'X\right)}} \frac{e^{\frac{1}{2\sigma^2} z' X \left(\sigma^2 D_{(\vartheta)}^{-1} + X'X\right)^{-1} X'z}}{e^{\frac{1}{2\sigma^2} z' X \left(\sigma^2 D_{(\delta)}^{-1} + X'X\right)^{-1} X'z}}.$$

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 $U, V \in \mathbb{R}^{n\times m}$) we have

$$\left(\frac{\rho_1}{\rho_0}\right)^{\frac{\|\vartheta\|_0 - \|\delta\|_0}{2}} \frac{\sqrt{\det\left(\sigma^2 D_{(\delta)}^{-1} + X'X\right)}}{\sqrt{\det\left(\sigma^2 D_{(\vartheta)}^{-1} + X'X\right)}} = \sqrt{\frac{\det\left(I_n + \frac{1}{\sigma^2} X D_{(\delta)} X'\right)}{\det\left(I_n + \frac{1}{\sigma^2} X D_{(\vartheta)} X'\right)}}.$$

By the Woodbury identity which states that for 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

$$X\left(\sigma^2 D_{(\delta)}^{-1} + X'X\right)^{-1} X' = \frac{1}{\sigma^2} X D_{(\delta)} X' - \frac{1}{\sigma^4} X D_{(\delta)} X' \left(I_n + \frac{1}{\sigma^2} X D_{(\delta)} X'\right)^{-1} X D_{(\delta)} X'$$
$$= I_n - \left(I_n + \frac{1}{\sigma^2} X D_{(\delta)} X'\right)^{-1}.$$

so that,

$$\frac{e^{\frac{1}{2\sigma^2} z' X \left(\sigma^2 D_{(\vartheta)}^{-1} + X'X\right)^{-1} X'z}}{e^{\frac{1}{2\sigma^2} z' X \left(\sigma^2 D_{(\delta)}^{-1} + X'X\right)^{-1} X'z}} = \frac{e^{\frac{1}{2\sigma^2} z' \left(I_n + \frac{1}{\sigma^2} X D_{(\delta)} X'\right)^{-1} z}}{e^{\frac{1}{2\sigma^2} z' \left(I_n + \frac{1}{\sigma^2} X D_{(\vartheta)} X'\right)^{-1} z}}.$$

We combine these developments to conclude that

$$\text{(SM2.2)} \qquad \frac{\Pi(\vartheta|z)}{\Pi(\delta|z)} = \left(\frac{1}{p^u}\right)^{\|\vartheta\|_0 - \|\delta\|_0} \sqrt{\frac{\det(L_\delta)}{\det(L_\vartheta)}} \frac{e^{\frac{1}{2\sigma^2} z' L_\delta^{-1} z}}{e^{\frac{1}{2\sigma^2} z' L_\vartheta^{-1} z}},$$

where, for $\delta \in \Delta$, we recall the definition $L_\delta \stackrel{\text{def}}{=} I_n + \frac{1}{\sigma^2} X D_{(\delta)} X'$. If $\vartheta \supseteq \delta$, setting $\tau \stackrel{\text{def}}{=} \frac{1}{\sigma^2}\left(\frac{1}{\rho_1} - \frac{1}{\rho_0}\right) < 1/(\sigma^2 \rho_1)$, it is easily seen that

$$L_\vartheta = L_\delta + \tau \sum_{j:\, \delta_j = 0, \vartheta_j = 1} X_j X_j'.$$

The determinant lemma then gives

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

And the Woodbury identity gives

$$L_\vartheta^{-1} = L_\delta^{-1} - \tau L_\delta^{-1} X_{(\vartheta-\delta)} \left(I_{\|\vartheta-\delta\|_0} + \tau X'_{(\vartheta-\delta)} L_\delta^{-1} X_{(\vartheta-\delta)}\right)^{-1} X'_{(\vartheta-\delta)} L_\delta^{-1}.$$

Combining the last two display in (SM2.2) yields the stated results. ∎

Lemma SM2.2. *Assume H2. Let ϱ and $\underline{\theta}_\star$ be as in Theorem 4.3. For $z \in \mathcal{E}_0$, we have*

$$\max_{\delta\in\Delta_s} \max_{1\le j\le p} \left|X'_j L_\delta^{-1} z - \theta_{\star j} X'_j L_\delta^{-1} X_j\right| \le \sqrt{\varrho n \log(p)}. \tag{SM2.3}$$

Furthermore, if $n \ge 4\varrho\log(p)/(\underline{\theta}_\star^2 \varpi_1^2)$, then

$$\min_{\delta\in\Delta_s} \min_{j:\ \delta_{\star j}=1} |X'_j L_\delta^{-1} z| \ge \frac{\varpi_1}{2} \underline{\theta}_\star n.$$

Proof. Set $V \stackrel{\text{def}}{=} (z - X\theta_\star)/\sigma$, so that

$$z = \sigma V + \sum_{k:\ \delta_{\star k}=1} \theta_{\star k} X_k,$$

and

$$X'_j L_\delta^{-1} z = \sigma X'_j L_\delta^{-1} V + \theta_{\star j} X'_j L_\delta^{-1} X_j + \sum_{k\ne j:\ \delta_{\star k}=1} \theta_{\star,k} X'_j L_\delta^{-1} X_k.$$

For $z \in \mathcal{E}_0$, $|X'_j L_\delta^{-1} V| \le \sqrt{c_0 n \log(p)}$. and for $k \ne j$, $|X'_j L_\delta^{-1} X_k| \le \mathcal{C}(s)\sqrt{n\log(p)}$. Hence

$$\max_{\delta\in\Delta_s} \max_{1\le j\le p} \left|X'_j L_\delta^{-1} z - \theta_{\star j} X'_j L_\delta^{-1} X_j\right| \le \left(\sigma\sqrt{c_0} + \mathcal{C}(s) \sum_{k:\ \delta_{\star k}=1} |\theta_{\star k}|\right) \sqrt{n\log(p)} \le \sqrt{\varrho n \log(p)}.$$

If $\delta_{\star j} = 1$, then $|X'_j L_\delta^{-1} z| \ge \frac{1}{2}|\theta_{\star,j}| X'_j L_\delta^{-1} X_j$, provided that we have $\sqrt{\varrho n \log(p)} \le \frac{1}{2}|\theta_{\star,j}| X'_j L_\delta^{-1} X_j$. Then using the definition of ϖ_1, we get

$$\min_{\delta\in\Delta_s} \min_{j:\ \delta_{\star j}=1} |X'_j L_\delta^{-1} z| \ge \frac{\varpi_1}{2} \underline{\theta}_\star n.$$

∎

Proof of Theorem 4.3. Fix $\varepsilon \in (0,1)$. We prove the result by applying (2.3) and Theorem 3.1. We recall that the initial distribution is taken as $\nu_0 = \Pi(\cdot|\delta^{(\mathrm{i})}, z)$, for some initial choice $\delta^{(\mathrm{i})} \in \Delta_s$. Let f_0 be the density of ν_0 with respect to $\Pi(\cdot|z)$. We then apply (2.3) to conclude that with $\varepsilon \in (0,1)$, we have

$$\left\|\nu_0 K^N - \Pi(\cdot|z)\right\|_{\mathrm{tv}} \le \|f_0\|_\infty (1 - \lambda(K))^N \le \varepsilon, \quad \text{for} \quad N \ge \frac{1}{\lambda(K)} \log\left(\frac{\|f_0\|_\infty}{\varepsilon}\right). \tag{SM2.4}$$

To bound the spectral gap we apply Theorem 3.1 with the choices $\zeta = 0$, $\mathsf{I} = \mathsf{I}_0 = \Delta_s$, and $\mathsf{B}_\delta = \mathbb{R}^p$, and with a graph on Δ_s constructed as follows: we put an edge between $\delta^{(1)}$ and $\delta^{(2)}$ if $\delta^{(1)} \supseteq \delta^{(2)}$, or $\delta^{(2)} \supseteq \delta^{(1)}$, and $\|\delta^{(2)} - \delta^{(1)}\|_0 = 1$ (in other words the models $\delta^{(1)}$ and $\delta^{(2)}$

differ only in one variable). Clearly (3.9) holds, since $\Pi(\Delta_s|z) = 1$. We then conclude from Theorem 3.1 that

$$\lambda(K) \geq \frac{\kappa}{1 + 8\mathsf{m}}. \tag{SM2.5}$$

To bound the constants κ and m we develop a similar argument as in [SM8]. Given $\delta \in \Delta_s$, we call $\min(\delta, \delta_\star)$ the skeleton of δ, and we let $\mathcal{S} \stackrel{\text{def}}{=} \{\min(\delta, \delta_\star),\ \delta \in \Delta_s\}$ be the set of all possible skeletons. Basically $\mathcal{S}$ is the set of submodels of the true model $\delta_\star$. Given $\delta \in \Delta_s$, we build our canonical path from δ to $\delta_\star$ as follows. First we build a path from δ to its skeleton (that is $\min(\delta, \delta_\star)$) by successively removing from the model δ the variables X_j for which $\delta_j = 1$ and $\delta_{\star j} = 0$, in reverse index ordering. Then we build a path from the skeleton to $\delta_\star$ by adding to the skeleton the variables X_j for which $\delta_j = 0$ and $\delta_{\star j} = 1$ in their index ordering. For example, if $p = 6$, $\delta_\star = (1, 1, 1, 0, 0)$ and $\delta = (0, 0, 1, 0, 1, 1)$, then our canonical path from δ to $\delta_\star$ is

$$(0, 0, 1, 0, 1, 1) \to (0, 0, 1, 0, 1, 0) \to (0, 0, 1, 0, 0, 0) \to (1, 0, 1, 0, 0, 0) \to (1, 1, 1, 0, 0, 0).$$

Given $\delta^{(1)}, \delta^{(2)} \in \Delta_s$, let $\delta^{(1,2)}$ be the element of Δ_s where the canonical path from $\delta^{(1)}$ to $\delta_\star$ and the canonical path from $\delta^{(2)}$ to $\delta_\star$ meet for the first time. Our canonical path $\gamma_{\delta^{(1)},\delta^{(2)}}$ between $\delta^{(1)}$ and $\delta^{(2)}$ is then defined as follows. Follow the canonical path from $\delta^{(1)}$ towards $\delta_\star$ until $\delta^{(1,2)}$, then reverse direction and follow the path from $\delta^{(1,2)}$ until $\delta^{(2)}$. For instance if $p = 6$, $\delta_\star = (1, 1, 1, 0, 0, 0)$ and $\delta^{(1)} = (0, 1, 0, 0, 1, 1)$, and $\delta^{(2)} = (1, 1, 0, 1, 1, 0)$, then $\delta^{(1,2)} = (1, 1, 0, 0, 0)$, and our chosen canonical path from $\delta^{(1)}$ to $\delta^{(2)}$ is

$$(0, 1, 0, 0, 1, 1) \to (0, 1, 0, 0, 1, 0) \to (0, 1, 0, 0, 0, 0) \to (1, 1, 0, 0, 0, 0) \to (1, 1, 0, 1, 0, 0) \to (1, 1, 0, 1, 1, 0).$$

We claim that for the canonical paths constructed above we have

$$\mathsf{m} \stackrel{\text{def}}{=} \max_{\delta \in \Delta_s} \sum_{\delta^{(1)},\delta^{(2)} \in \Delta_s:\ \gamma_{\delta^{(1)},\delta^{(2)}} \ni \delta} |\gamma_{\delta^{(1)},\delta^{(2)}}| \frac{\pi(\delta^{(1)}|z)\pi(\delta^{(2)}|z)}{\pi(\delta|z)} \leq 16s, \tag{SM2.6}$$

and

$$\kappa \stackrel{\text{def}}{=} \min_{\delta^{(1)} \sim \delta^{(2)}} \int_{\mathbb{R}^p} \min\left(\Pi(\theta|\delta^{(1)}, z), \Pi(\theta|\delta^{(2)}, z)\right) \mathrm{d}\theta \geq C_0 e^{-2\rho_0 \|\theta_\star\|_\infty^2}, \tag{SM2.7}$$

for some absolute constant C_0, where the minimum is taken over all connected pairs of nodes $\delta^{(1)}, \delta^{(2)}$. Furthermore, we claim that we can bound the variance of the initial density and get

$$\log\left(\frac{\|f\|_\infty}{\varepsilon}\right) \leq C_0 \left(\log\left(\frac{1}{\varepsilon}\right) + s_\star \|\theta_\star\|_\infty^2 n\right), \tag{SM2.8}$$

for some absolute constant C_0. (SM2.6) and (SM2.7) shows that

$$\lambda(K) \geq C_0 s^{-1} e^{-2\rho_0 \|\theta_\star\|_\infty^2}, \tag{SM2.9}$$

for some absolute constant C_0. We put (SM2.9) together with (SM2.8) and (SM2.4) to reach the stated conclusion. The remaining of the proof consists in establishing the claims (SM2.6), (SM2.7), and (SM2.8).

Proof of Equation (SM2.6). For $\delta^{(1)}, \delta^{(2)} \in \Delta_s$, we will use the obvious bound

$$|\gamma_{\delta^{(1)},\delta^{(2)}}| \leq 2s.$$

Given $\delta \in \Delta_s$, we denote $\Lambda(\delta)$ the set of all $\delta^{(1)} \in \Delta_s$ such that the canonical path from $\delta^{(1)}$ to $\delta_\star$ goes through δ. If a path $\gamma_{\delta^{(1)},\delta^{(2)}}$ goes through δ, then $\delta^{(1)} \in \Lambda(\delta)$ or $\delta^{(2)} \in \Lambda(\delta)$. Using this we can bound m as

$$\mathsf{m} \leq 4s \max_{\delta\in\Delta_s} \sum_{\delta^{(1)}\in\Lambda(\delta)} \sum_{\delta^{(2)}\in\Delta_s} \frac{\pi(\delta^{(1)}|z)\pi(\delta^{(2)}|z)}{\pi(\delta|z)} \leq 4s \max_{\delta\in\Delta_s} \sum_{\delta^{(1)}\in\Lambda(\delta)} \frac{\pi(\delta^{(1)}|z)}{\pi(\delta|z)}. \tag{SM2.10}$$

Let $\mathcal{S} \stackrel{\text{def}}{=} \{\min(\delta,\delta_\star),\ \delta \in \Delta_s\}$ be the set of all possible skeletons. Take $\delta^{(1)} \in \Lambda(\delta)$. We will distinguish whether $\delta \in \mathcal{S}$ or not. Suppose $\delta \notin \mathcal{S}$. Therefore, traveling the canonical path from $\delta^{(1)}$ toward $\delta_\star$ we arrive at δ by removing only non-significant variables. Therefore, assuming that $\|\delta^{(1)}\|_0 = \|\delta\|_0 + \ell$, and using (SM2.1), and H2-(3), we have

$$\frac{\pi(\delta^{(1)}|z)}{\pi(\delta|z)} \leq \frac{1}{p^{u\ell}} \exp\left(\frac{\tau}{2\sigma^2(1+n\tau\varpi_{s_0})} \sum_{j:\ \delta_j^{(1)}=1,\delta_j=0} (X_j' L_\delta^{-1} z)^2\right) \leq \frac{e^{\frac{\ell\bar{Q}_0}{n\varpi_{s_0}}}}{p^{u\ell}}, \tag{SM2.11}$$

where $\bar{Q}_0 = \max_{j:\ \delta_j^{(1)}=1,\delta_j=0} (X_j' L_\delta^{-1} z)^2$. From Lemma SM2.2, we get $\bar{Q}_0 \leq \varrho n \log(p)$. Using this and the trivial inequality $\binom{p}{\ell} \leq p^\ell$, it follows that

$$\sum_{\delta^{(1)}\in\Lambda(\delta)} \frac{\pi(\delta^{(1)}|z)}{\pi(\delta|z)} \leq \sum_{\ell=0}^{s-\|\delta\|_0} \sum_{\delta^{(1)}\in\Lambda(\delta):\ \|\delta^{(1)}\|_0=\|\delta\|_0+\ell} \frac{\pi(\delta^{(1)}|z)}{\pi(\delta|z)} \leq \sum_{\ell=0}^{s} \left(\frac{p^{\frac{\varrho}{2\sigma^2\varpi_{s_0}}}}{p^{u-1}}\right)^\ell \leq 2,$$

under the assumption that $\sigma^2 u\varpi_{s_0} \geq \varrho$, and $u > 4$. Suppose now that $\delta \in \mathcal{S}$. Then $\Lambda(\delta)$ is comprised of the elements of Δ_s whose skeletons are subsets of δ. Hence

$$\sum_{\delta^{(1)}\in\Lambda(\delta)} \frac{\pi(\delta^{(1)}|z)}{\pi(\delta|z)} = \sum_{\delta_0\in\mathcal{S}:\delta\supseteq\delta_0} \frac{\pi(\delta_0|z)}{\pi(\delta|z)} \sum_{\delta^{(1)}\in\Lambda(\delta):\ \min(\delta^{(1)},\delta_\star)=\delta_0} \frac{\pi(\delta^{(1)}|z)}{\pi(\delta_0|z)}.$$

The inner summation can be upper bounded by 2 as above. If $\delta \supseteq \delta_0$ and $\|\delta\|_0 = \|\delta_0\|_0 + r$, we apply (SM2.1) again and get,

$$\frac{\pi(\delta_0|z)}{\pi(\delta|z)} \leq \left(p^u \sqrt{1+\frac{ns_\star}{\sigma^2\rho_1}}\, e^{-\frac{\tau\bar{Q}_3}{2\sigma^2(1+\tau s_\star n)}}\right)^r \leq \left(p^{u+a} e^{-\frac{\bar{Q}_3}{4\sigma^2 s_\star n}}\right)^r,$$

where we use H2-(3) to obtain $\tau/(1+\tau s_\star n) \geq 1/(2s_\star n)$, and $\sqrt{1+\frac{ns_\star}{\sigma^2\rho_1}} \leq p^a$, and where $\bar{Q}_3 \stackrel{\text{def}}{=} \min_{j:\ \delta_{0j}=0,\delta_{\star j}=1} \left(X_j' L_{\delta_0}^{-1} z\right)^2$. From Lemma SM2.2 we get $\bar{Q}_3 \geq \frac{\underline{\theta}_\star^2}{4}\varpi_1^2 n^2$, under the sample condition $n \geq 4\varrho\log(p)/(\underline{\theta}_\star^2\varpi_1^2)$ which is implied by (4.11). We conclude that

$$\max_{\delta\in\mathcal{S}} \sum_{\delta^{(1)}\in\Lambda(\delta)} \frac{\pi(\delta^{(1)}|z)}{\pi(\delta|z)} \leq 2 \sum_{\delta_0\in\mathcal{S}:\delta\supseteq\delta_0} \frac{\pi(\delta_0|z)}{\pi(\delta|z)} \leq 2\sum_{r=0}^{s_\star} s_\star^r \left(p^{u+a} e^{-\frac{\underline{\theta}_\star^2\varpi_1^2 n}{4\sigma^2 s_\star}}\right)^r \leq 4,$$

using the sample size condition in (4.11). This proves the claim (SM2.6). ∎

Proof of Equation (SM2.7). Fix $\delta^{(1)}, \delta^{(2)} \in \Delta_s$, such that $\delta^{(1)} \supseteq \delta^{(2)}$, or $\delta^{(2)} \subseteq \delta^{(1)}$, and $\|\delta^{(2)} - \delta^{(1)}\|_0 = 1$. Without any loss of generality, suppose that $\delta^{(2)} \supseteq \delta^{(1)}$, and their difference occurs on component j: $\delta_j^{(2)} = 1$, while $\delta_j^{(1)} = 0$. Then for all $\theta \in \mathbb{R}^p$, we have

$$\frac{\Pi(\theta|\delta^{(1)}, z)}{\Pi(\theta|\delta^{(2)}, z)} = \left(\frac{\int_{\mathbb{R}^p} e^{-\frac{1}{2\sigma^2}\|z-X\theta\|_2^2 - \frac{1}{2}\theta' D^{-1}_{(\delta^{(2)})}\theta} \mathrm{d}\theta}{\int_{\mathbb{R}^p} e^{-\frac{1}{2\sigma^2}\|z-X\theta\|_2^2 - \frac{1}{2}\theta' D^{-1}_{(\delta^{(1)})}\theta} \mathrm{d}\theta}\right) e^{-(\rho_0-\rho_1)\frac{\theta_j^2}{2}}.$$

Let A denote the ratio of integrals in the last display. We can then write

$$\int_{\mathbb{R}^p} \min\left(\Pi(\theta|\delta^{(1)}, z), \Pi(\theta|\delta^{(2)}, z)\right) \mathrm{d}\theta = \int_{\mathbb{R}} \min\left(1, Ae^{-(\rho_0-\rho_1)\frac{\theta_j^2}{2}}\right) \Pi(\theta_j|\delta^{(2)}, z)\mathrm{d}\theta_j.$$

Recall from (4.3) that the j marginal under $\Pi(\theta_j|\delta^{(2)}, z)$ is the Gaussian distribution $\mathbf{N}(\mu_j, \sigma_j^2)$, where

$$\sigma_j = \sigma\sqrt{e_j'\Sigma_{\delta^{(2)}} e_j}, \quad \text{and} \quad \mu_j = e_j'\Sigma_{\delta^{(2)}} X' z, \quad 1 \leq j \leq p,$$

and where e_j denotes the j-th unit vector. Hence, for $Z \sim \mathbf{N}(0, 1)$,

$$\begin{aligned}\text{(SM2.12)} \quad \int_{\mathbb{R}^p} \min\left(\Pi(\theta|\delta^{(1)}, z), \Pi(\theta|\delta^{(2)}, z)\right) \mathrm{d}\theta &= \mathbb{E}\left[\min\left(1, Ae^{-\frac{(\rho_0-\rho_1)}{2}(\mu_j+\sigma_j Z)^2}\right)\right] \\ &\geq \frac{1}{2}\min\left(1, Ae^{-\frac{(\rho_0-\rho_1)}{2}(|\mu_j|+\sigma_j)^2}\right) \geq \frac{1}{2}\min\left(1, Ae^{-\rho_0(\mu_j^2+\sigma_j^2)}\right),\end{aligned}$$

using the fact that for any nonnegative function f, $\mathbb{E}(f(Z)) \geq \mathbb{P}(|Z| \leq 1)\min_{z:|z|\leq 1} f(z)$. By matrix block inversion, we work out σ_j^2 to

$$\text{(SM2.13)} \quad \sigma_j^2 = \frac{\sigma^2}{\sigma^2\rho_1 + X_j'\left(I_n + \frac{1}{\sigma^2}X_{-j}D_{(\delta^{(2)},j)}X_{-j}'\right)^{-1} X_j} = \frac{\sigma^2}{\sigma^2\rho_1 + X_j' L^{-1}_{\delta^{(1)}_{-j}} X_j} \leq \frac{\sigma^2}{\varpi_1 n},$$

where $D_{(\delta^{(2)},j)} = D_{(\delta^{(1)},j)}$ is the $(p-1)$-dimensional matrix obtained by removing the j-th row and the j-th column of $D_{(\delta^{(2)})}$, and $L_{\delta^{(1)}_{-j}} = I_n + \frac{1}{\sigma^2}X_{-j}D_{(\delta^{(1)},j)}X_{-j}'$. By block inversion the mean μ_j can be written as

(SM2.14)

$$\mu_j = e_1\begin{pmatrix} X_j'X_j + \sigma^2\rho_1 & X_j'X_{-j} \\ X_{-j}'X_j & X_{-j}'X_{-j} + \sigma^2 D^{-1}_{(\delta^{(2)},j)}\end{pmatrix}^{-1}\begin{pmatrix} X_j'z \\ X_{-j}'z\end{pmatrix} = \frac{X_j' L^{-1}_{\delta^{(1)}_{-j}} z}{\sigma^2\rho_1 + X_j' L^{-1}_{\delta^{(1)}_{-j}} X_j}.$$

Consider first the case where j is such that $\delta_{\star,j} = 0$. Note that $X_j' L^{-1}_{\delta^{(1)}_{-j}} X_j \geq X_j' L^{-1}_{\delta^{(1)}} X_j \geq n\varpi_1$. Therefore, and using Lemma SM2.2, and $z \in \mathcal{E}_0$, we obtain

$$\sigma_j^2 + \mu_j^2 \leq \frac{\sigma^2}{n\varpi_1} + \frac{\varrho\log(p)}{n\varpi_1^2} \leq \frac{2\varrho\log(p)}{n\varpi_1^2}.$$

Consider now the case where $\delta_{\star,j} = 1$. Then using Lemma SM2.2, we have

$$\begin{aligned}
\sigma_j^2 + \mu_j^2 &\leq \sigma_j^2 + \frac{1}{(X_j' L^{-1}_{\delta^{(1)}_{-j}} X_j)^2}\left(\theta_{\star j} X_j' L^{-1}_{\delta^{(1)}_{-j}} X_j + \sqrt{\varrho n \log(p)}\right)^2 \\
&\leq \sigma_j^2 + \theta_{\star j}^2 \left(1 + \frac{\sqrt{\varrho n \log(p)}}{\theta_{\star j} X_j' L^{-1}_{\delta^{(1)}_{-j}} X_j}\right)^2 \\
&\leq \sigma_j^2 + \theta_{\star j}^2 \left(1 + \frac{\sqrt{\varrho n \log(p)}}{\theta_{\star j} n \varpi_1}\right)^2 \\
&\leq 2|\theta_{\star j}|^2,
\end{aligned}$$

under the sample size condition (4.11). On the other hand, using (SM2.1), the ratio of integrals A gives

$$A = \sqrt{\frac{\rho_0}{\rho_1}} \frac{1}{\sqrt{1 + \tau X_j' L^{-1}_{\delta^{(1)}} X_j}} \exp\left(\frac{1}{2\sigma^2} \frac{\tau (X_j' L^{-1}_{\delta^{(1)}} z)^2}{1 + \tau X_j' L^{-1}_{\delta^{(1)}} X_j}\right),$$

where we recall that $\tau = (\rho_1^{-1} - \rho_0^{-1})/\sigma^2$. Note that if $\delta_{\star j} = 0$, the term inside the exponential in this last expression of A grows like $\varrho \log(p)/\varpi_1$ which is not fast enough to face off with the term $-\rho_0(\mu_j^2 + \sigma_j^2)$. Hence we use instead the trivial lower bound $A \geq 1$ together with the upper bounds on μ_j and σ_j^2 obtained above and (SM2.12) to conclude that

$$\text{(SM2.15)} \quad \int_{\mathbb{R}^p} \min\left(\Pi(\theta|\delta^{(1)}, z), \Pi(\theta|\delta^{(2)}, z)\right) \mathrm{d}\theta \geq \frac{1}{2} e^{-\rho_0(\mu_j^2 + \sigma_j^2)} \geq \frac{1}{2}\exp\left(-\frac{2\rho_0}{n}\frac{\varrho \log(p)}{\varpi_1^2}\right).$$

However if $\delta_{\star j} = 1$, By Lemma SM2.2, and under the sample size condition (4.11) we have

$$(X_j' L^{-1}_{\delta^{(1)}} z)^2 \geq \frac{\theta_{\star j}^2}{2} (X_j' L^{-1}_{\delta^{(1)}} X_j)^2.$$

Noting that $1 \leq \tau X_j' L^{-1}_{\delta^{(1)}} X_j$, we deduce that

$$A \geq \sqrt{\frac{\rho_0}{\rho_1}} \frac{1}{\sqrt{1 + \frac{n}{\sigma^2 \rho_1}}} e^{\frac{\theta_{\star j}^2 X_j' L^{-1}_{\delta^{(1)}} X_j}{4\sigma^2}} \geq \sqrt{\frac{\sigma^2 \rho_0}{\sigma^2 \rho_1 + n}} e^{\frac{n \varpi_1 \theta_{\star j}^2}{4\sigma^2}}.$$

It follows in this case that

$$\begin{aligned}
\text{(SM2.16)} \quad \int_{\mathbb{R}^p} \min\left(\Pi(\theta|\delta^{(1)}, z), \Pi(\theta|\delta^{(2)}, z)\right) \mathrm{d}\theta &\geq \frac{1}{2} \min\left(1, A e^{-2\rho_0 \theta_{\star j}^2}\right) \\
&\geq \frac{1}{2} \min\left(1, \sqrt{\frac{\sigma^2 \rho_0}{\sigma^2 \rho_1 + n}}\right) \min\left(1, e^{-2\rho_0 \theta_{\star j}^2 \left(1 - \frac{n \varpi_1}{8\rho_0}\right)}\right) \geq C_0 e^{-2\rho_0 \|\theta_\star\|_\infty^2},
\end{aligned}$$

for some absolute constant C_0, where we have used the fact that $\min(1, ab) \geq \min(1, a)\min(1, b)$ valid for all nonnegative numbers a, b, c. We combine (SM2.15) and (SM2.16) to obtain (SM2.7).

Proof of Equation (SM2.8). Since $\Pi(\theta|z) = \sum_{\vartheta} \Pi(\vartheta|z)\Pi(\theta|\vartheta, z) \geq \Pi(\delta^{(\mathrm{i})}|z)\Pi(\theta|\delta^{(\mathrm{i})}, z)$, we have

$$f_0(\theta) = \frac{\Pi(\theta|\delta^{(\mathrm{i})}, z)}{\Pi(\theta|z)} \leq \frac{1}{\Pi(\delta^{(\mathrm{i})}|z)} = \frac{1}{\Pi(\delta_\star|z)} \frac{\Pi(\delta_\star|z)}{\Pi(\delta_\star^{(\mathrm{i})}|z)} \frac{\Pi(\delta_\star^{(\mathrm{i})}|z)}{\Pi(\delta^{(\mathrm{i})}|z)},$$

where $\delta_\star^{(\mathrm{i})} \stackrel{\text{def}}{=} \min(\delta^{(\mathrm{i})}, \delta_\star)$. We apply (SM2.1) twice (to each ratio), and use H??, to get

$$\frac{\Pi(\delta_\star|z)}{\Pi(\delta_\star^{(\mathrm{i})}|z)} \frac{\Pi(\delta_\star^{(\mathrm{i})}|z)}{\Pi(\delta|z)} \leq p^{u(\|\delta\|_0 - \|\delta_\star\|_0)} \sqrt{\det\left(I_{\|\delta\|_0 - \|\delta_\star^{(\mathrm{i})}\|_0} + \tau X'_{(\delta-\delta_\star^{(\mathrm{i})})} L^{-1}_{\delta_\star^{(\mathrm{i})}} X_{(\delta-\delta_\star^{(\mathrm{i})})}\right)}$$

$$\times\, e^{\frac{\tau}{2\sigma^2} z' L^{-1}_{\delta_\star^{(\mathrm{i})}} X_{(\delta_\star-\delta_\star^{(\mathrm{i})})} \left(I_{\|\delta_\star-\delta_\star^{(\mathrm{i})}\|_0} + \tau X'_{(\delta_\star-\delta_\star^{(\mathrm{i})})} L^{-1}_{\delta_\star^{(\mathrm{i})}} X_{(\delta_\star-\delta_\star^{(\mathrm{i})})}\right)^{-1} X'_{(\delta_\star-\delta_\star^{(\mathrm{i})})} L^{-1}_{\delta_\star^{(\mathrm{i})}} z}$$

$$\leq p^{u(\|\delta\|_0 - s_\star)} \left(1 + \frac{n\|\delta\|_0}{\sigma^2\rho_1}\right)^{\frac{\|\delta\|_0}{2}} e^{\frac{1}{2\sigma^2 n\varpi_1} \|X'_{(\delta_\star-\delta_\star^{(\mathrm{i})})} L^{-1}_{\delta_\star^{(\mathrm{i})}} z\|_2^2}.$$

Under the assumption $p^{us_\star}\Pi(\delta_\star|z) \geq 1$ (H2-(1)), and since $\|\delta\|_0 \leq s$, we conclude that

$$\|f_0\|_\infty \leq p^{us} \left(1 + \frac{ns}{\sigma^2\rho_1}\right)^{\frac{s}{2}} e^{\frac{s_\star \bar{Q}_1}{2\sigma^2 n\varpi_1}} \leq p^{(u+a)s} e^{\frac{s_\star \bar{Q}_1}{2\sigma^2 n\varpi_1}}, \tag{SM2.17}$$

where the second inequality uses (4.9), and where $\bar{Q}_1 = \max_{j:\delta_{\star,j}=1} (X_j' L^{-1}_{\delta_\star^{(\mathrm{i})}} z)^2$. From Lemma SM2.2, we get $\bar{Q}_1 \leq 4n^2\|\theta_\star\|_\infty^2$, using the sample size condition (4.11). (SM2.17) then becomes

$$\|f_0\|_\infty \leq p^{(u+a)s} e^{\frac{2s_\star\|\theta_\star\|_\infty^2 n}{\sigma^2\varpi_1}}.$$

The claim follows by taking the log. ■

SM3. Proof of Theorem 4.4. The proof is very similar to the proof of Theorem 4.3, but here we use the approximate spectral gap with $\mathsf{I} = \Delta$, $\mathsf{I}_0 = \mathcal{D}_k$ for some k specified below, for some ζ specified below, and $\mathsf{B}_\delta = \mathbb{R}^p$. First we apply (2.7) (with $\lambda_\zeta(K^\star K)$ replaced by $\lambda_\zeta(K)$, since K is positive) to conclude that for any $\zeta \in (0,1]$, we have

$$\left\|\nu_0 K^N - \Pi(\cdot|z)\right\|_{\mathrm{tv}} \leq \|f_0\|_\infty (1 - \min(1, \lambda_\zeta(K)))^{N/2} + \|f_0\|_\infty \sqrt{\zeta} \leq \varepsilon + \|f_0\|_\infty\sqrt{\zeta},$$

provided that N satisfies $N \geq (\min(1, \lambda_\zeta(K)))^{-1} \log\left(\frac{\|f_0\|_\infty^2}{\varepsilon^2}\right)$.

We bound the uniform norm $\|f_0\|_\infty$ as in (SM2.17). Set $\mathsf{FP} \stackrel{\text{def}}{=} \|\delta^{(\mathrm{i})}\|_0 - \|\delta_\star\|_0$. Noting here that the skeleton of $\delta^{(\mathrm{i})}$ is $\delta_\star$, so we get the simpler bound

$$\|f_0\|_\infty \leq 2\left(p^u \sqrt{1 + \frac{n\mathsf{FP}}{\sigma^2\rho_1}}\right)^{\mathsf{FP}} \leq 2p^{(u+a)\mathsf{FP}}.$$

In view of this bound, we set

$$\zeta = \frac{\varepsilon^2}{4} p^{-2(u+a)\mathsf{FP}}, \tag{SM3.1}$$

which gives $\|f_0\|_\infty\sqrt{\zeta} \leq \varepsilon$, and we conclude that with ζ as in (SM3.1), we have

$$\|\nu_0 K^N - \Pi(\cdot|z)\|_{\mathrm{tv}} \leq 2\varepsilon, \quad \text{for} \quad N \geq \frac{C_0}{\min(1,\lambda_\zeta(K))}\left(\log\left(\frac{1}{\varepsilon^2}\right) + (u+a)\mathsf{FP}\log(p)\right), \tag{SM3.2}$$

for some absolute constant C_0. We lower bound the approximate spectral gap via Theorem 3.1, and using the same approach as in Theorem 4.3. First note that for $z \in \mathcal{E}$, ζ as in (SM3.1), and let $k \geq 1$, be the first integer such that

$$\frac{10}{\zeta}\left(1 - \Pi(\mathcal{D}_k|z)\right) \leq \frac{40}{\varepsilon^2} p^{2(u+a)\mathsf{FP}} \frac{1}{p^{\frac{u(k+1)}{2}}} \leq 1.$$

Under condition (4.13), this choice of k also satisfies $\mathcal{D}_k \subseteq \Delta_{s_0}$, where s_0 is as in H2. We then apply Theorem 3.1 with the choices $\mathsf{I} = \Delta$, $\mathsf{I}_0 = \mathcal{D}_k$ endowed with the same graph as in proof of Theorem 4.3, and $\mathsf{B}_\delta = \mathbb{R}^p$. Given our particular choice of k, (3.9) holds. We then conclude from Theorem 3.1 that

$$\lambda_\zeta(K) \geq \frac{\kappa}{1 + 8\mathsf{m}}, \tag{SM3.3}$$

where κ and m are defined using $\mathcal{D}_k$. We bound these terms as in Theorem 4.3 with some important simplifications due the facts that all models here belong to $\mathcal{D}_k$. In particular

$$\mathsf{m} \leq 16(s_\star + k) \leq C_0\|\delta^{(\mathrm{i})}\|_0, \tag{SM3.4}$$

for some absolute constant C_0. Similarly, the lower bound on κ also simplifies. Because $\delta^{(1)}$ and $\delta^{(2)}$ can differ only at a component j such that $\delta_{\star j} = 0$ (a non-important variable), we see that only the lower bound (SM2.15) applies. Hence κ can be taken as

$$\kappa = \frac{1}{2} p^{-\frac{2c_1\varrho}{\varpi_1}}. \tag{SM3.5}$$

The theorem follows from the same calculations as in the proof of Theorem 4.3. ■

SM4. Some technical results. We make use of the following standard Gaussian deviation bound.

Lemma SM4.1. *Let $Z \sim \boldsymbol{N}(0, I_m)$, and $u_1,\ldots,u_N$ be vectors of $\mathbb{R}^m$. Then for all $x \geq 0$,*

$$\mathbb{P}\left[\max_{1\leq j\leq N} |\langle u_j, Z\rangle| > \max_{1\leq j\leq N}\|u_j\|_2\sqrt{2(x+\log(N))}\right] \leq \frac{2}{e^x}.$$

Lemma SM4.2. *Suppose that $X \in \mathbb{R}^{n\times p}$ is a random matrix with i.i.d. standard Normal entries. Given an integer s, and positive constants σ, γ and ρ, set*

$$\mathcal{C}_0 \stackrel{\text{def}}{=} \max_{\delta\in\Delta:\ \|\delta\|_0\leq s}\ \max_{i\neq j,\ \delta_j=0} \left| X_j'\left(I_n + \frac{1}{\sigma^2\rho_1}X_\delta X_\delta' + \frac{1}{\sigma^2\rho_0}X_{\delta^c}X_{\delta^c}'\right)X_i\right|.$$

Then there exist some universal finite constants c_0, a, A such that for $n \geq As^2\log(p)$, the following two statements hold with probability at least $1 - \frac{a}{p}$: for $\rho_0^{-1} > 0$ taken small enough and

$$\sigma^2 s\rho_1 \leq c_0\sqrt{n\log(p)}, \tag{SM4.1}$$

it holds that

(SM4.2) $\mathcal{C}_0 \leq 2c_0\sqrt{n\log(p)}$, *and*

$$\min_{\delta:\ \|\delta\|_0\leq s}\ \inf\left\{\frac{u'(X'_{\delta^c}L_\delta^{-1}X_{\delta^c})u}{n\|u\|_2^2},\ u\in\mathbb{R}^{p-s},\ 0<\|\mathit{supp}(u)\|_0\leq s\right\}\geq\frac{1}{32}.$$

Proof. For a matrix $M\in\mathbb{R}^{n\times p}$ we set

$$v(M,s)\stackrel{\text{def}}{=}\inf\left\{\frac{u'(M'M)u}{n\|u\|_2^2}\ u\neq 0, \|u\|_0\leq s\right\},$$

and for $\kappa_0 = 1/64$ and $c_0 = 8$, we define

$$\mathcal{E}\stackrel{\text{def}}{=}\left\{M\in\mathbb{R}^{n\times p}:\ v(M,s)\geq\kappa_0,\ \ \max_{1\leq j\leq p}\|M_j\|_2\leq 2\sqrt{n},\right.$$
$$\left.\min_{1\leq j\leq p}\|M_j\|_2\geq\sqrt{\frac{n}{2}},\quad\text{and}\quad\max_{j\neq k}|\langle M_j,M_k\rangle|\leq c_0\sqrt{n\log(p)}\right\}.$$

By Theorem 1 of [SM6], Lemma 1-(4.2) of [SM4], and standard Gaussian deviation bounds, we can find universal constants a, A, such that for $n\geq As\log(p)$, we have $\mathbb{P}(X\notin\mathcal{E})\leq\frac{a}{p}$. So to obtained the statement of the lemma, it suffices to consider some arbitrary element $X\in\mathcal{E}$ and show that (SM4.2) holds.

Fix $\delta\in\Delta$ such that $\|\delta\|_0\leq s$. We set $M_\delta\stackrel{\text{def}}{=}I_n+\frac{1}{\sigma^2\rho_1}X_\delta X'_\delta$, so that $L_\delta = M_\delta+\frac{1}{\sigma^2\rho_0}X_{\delta^c}X'_{\delta^c}$. The Woodbury identity gives

(SM4.3)
$$X'_jL_\delta^{-1}X_k = X'_jM_\delta^{-1}X_k - \frac{1}{\sigma^2\rho_0}X'_jM_\delta^{-1}X_{\delta^c}\left(I_{\|\delta^c\|_0}+\frac{1}{\sigma^2\rho_0}X'_{\delta^c}M_\delta^{-1}X_{\delta^c}\right)^{-1}X'_{\delta^c}M_\delta^{-1}X_k.$$

If $C_1 = \max_\ell X'_\ell M_\delta^{-1}X_\ell$, and $C_0 = \max_{\ell\neq j,\ \delta_j=0}|X'_jM_\delta^{-1}X_\ell|$, then we deduce easily from (SM4.3) that for all $j\neq k$ such that $\delta_j = 0$,

$$|X'_jL_\delta^{-1}X_k|\leq C_0+\frac{1}{\sigma^2\rho_0}\left(C_1^2+pC_0^2\right). \tag{SM4.4}$$

In order to proceed, we need to bound the term $X_jM_\delta^{-1}X_k$. Easily, for $X\in\mathcal{E}$, we have

$$X'_jM_\delta^{-1}X_j\leq\|X_j\|_2^2\leq 4n.$$

Another application of the Woodbury identity gives

$$M_\delta^{-1} = I_n-\frac{1}{\sigma^2\rho_1}X_\delta\left(I_{\|\delta\|_0}+\frac{1}{\sigma^2\rho}X'_\delta X_\delta\right)^{-1}X'_\delta. \tag{SM4.5}$$

Therefore, for $k\neq j$

$$X'_jM_\delta^{-1}X_k = X'_jX_k-\frac{1}{\sigma^2\rho_1}X'_jX_\delta\left(I_{\|\delta\|_0}+\frac{1}{\sigma^2\rho}X'_\delta X_\delta\right)^{-1}X'_\delta X_k.$$

Using $X \in \mathcal{E}$, we educe for $j \neq k$, and $\delta_j = 0$,

$$\frac{1}{\sigma^2\rho_1}\left|X_j' X_\delta \left(I_{\|\delta\|_0} + \frac{1}{\sigma^2\rho} X_\delta' X_\delta\right)^{-1} X_\delta' X_k\right| \leq \frac{1}{\kappa_0 n}\|X_\delta' X_k\|_2 \|X_\delta' X_j\|_2$$
$$\leq \frac{c_0^2 s\log(p) + c_0\sqrt{s\log(p)}}{\kappa_0} \leq c_0\sqrt{n\log(p)},$$

for $n \geq As^2\log(p)$, for some constant A. It follows that

$$|X_j' M_\delta^{-1} X_k| \leq 2c_0\sqrt{n\log(p)}.$$

We combine this with (SM4.4) to obtain that for $j \neq k$ such that $\delta_j = 0$,

$$|X_j' L_\delta^{-1} X_k| \leq 3c_0\sqrt{n\log(p)}\left(1 + \frac{1}{\sigma^2\rho_0} pc_0\sqrt{n\log(p)}\right) + 16\frac{1}{\sigma^2\rho_0} n^2 \leq 8c_0\sqrt{n\log(p)}, \tag{SM4.6}$$

for ρ_0 large enough. (SM4.6) says that $\mathcal{C}_0 \leq 8c_0\sqrt{n\log(p)}$, for $X \in \mathcal{E}$, as claimed.

For j such that $\delta_j = 0$, (SM4.5) gives

$$X_j' M_\delta^{-1} X_j = \|X_j\|_2^2 - \frac{1}{\sigma^2\rho_1} X_j' X_\delta\left(I_{\|\delta\|_0} + \frac{1}{\sigma^2\rho_1} X_\delta' X_\delta\right)^{-1} X_\delta' X_j$$
$$\geq \|X_j\|_2^2 - \frac{\|X_\delta' X_j\|_2^2}{n\kappa_0}$$
$$\geq \frac{n}{4}, \tag{SM4.7}$$

since $n \geq As\log(p)$, and by taking A large enough ($A \geq 4c_0^2/\kappa_0$). Equation (**??**) then yields

$$X_j' L_\delta^{-1} X_j \geq X_j' M_\delta^{-1} X_j - \frac{1}{\sigma^2\rho_0}\|X_{\delta^c}' M_\delta^{-1} X_j\|_2^2$$
$$= X_j' M_\delta^{-1} X_j - \frac{1}{\sigma^2\rho_0}\left[(X_j' M_\delta^{-1} X_j)^2 + \sum_{k:\ \delta_k = 0, k\neq j}(X_j' M_\delta^{-1} X_k)^2\right].$$

For $2\rho_0^{-1} \leq \sigma^2$, it follows that

$$X_j' L_\delta^{-1} X_j \geq \frac{n}{8} - \frac{1}{\sigma^2\rho_0}(p - \|\delta\|_0)\left(4c_0^2 n\log(p)\right),$$

which together with (SM4.6) and (SM4.1) implies that for any $u \in \mathbb{R}^p$ such that $\delta^c \supseteq \mathsf{supp}(u)$, and $\|\mathsf{supp}(u)\|_0 \leq s$, we have

$$u' X_{\delta^c}' L_\delta^{-1} X_{\delta^c} u \geq \frac{n}{32}\|u\|_2^2, \qquad \blacksquare$$

as claimed.

REFERENCES

[1] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin, *A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case*, Electronic Communications in Probability, 13 (2008).

[2] D. Bakry, I. Gentil, and M. LeDoux, *Analysis and Geometry of Markov Diffusion Operators*, Springer, 2013.

[3] S. G. Bobkov, *Isoperimetric and analytic inequalities for log-concave probability measures*, Ann. Probab., 27 (1999), pp. 1903–1921.

[4] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist., 28 (2000), pp. 1302–1338, https://doi.org/10.1214/aos/1015957395.

[5] M. Raginsky, A. Rakhlin, and M. Telgarsky, *Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis*, in Proceedings of the 2017 Conference on Learning Theory 65, Proceedings of Machine Learning Research, PMLR, 07–10 Jul 2017, pp. 1674–1703.

[6] G. Raskutti, M. J. Wainwright, and B. Yu, *Restricted eigenvalue properties for correlated gaussian designs*, J. Mach. Learn. Res., 11 (2010), pp. 2241–2259.

[7] M. Welling and Y. W. Teh, *Bayesian learning via stochastic gradient Langevin dynamics*, in Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11, USA, 2011, Omnipress, pp. 681–688, http://dl.acm.org/citation.cfm?id=3104482.3104568.

[8] Y. Yang, M. J. Wainwright, and M. I. Jordan, *On the computational complexity of high-dimensional Bayesian variable selection*, Ann. Statist., 44 (2016), pp. 2497–2532.