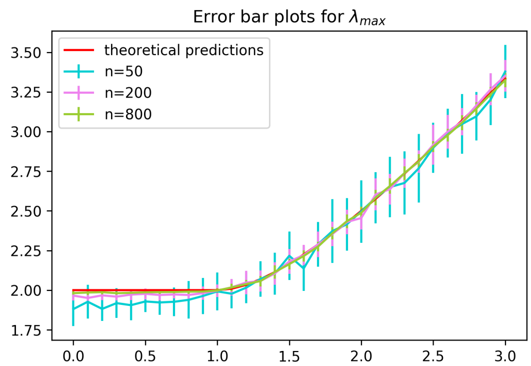
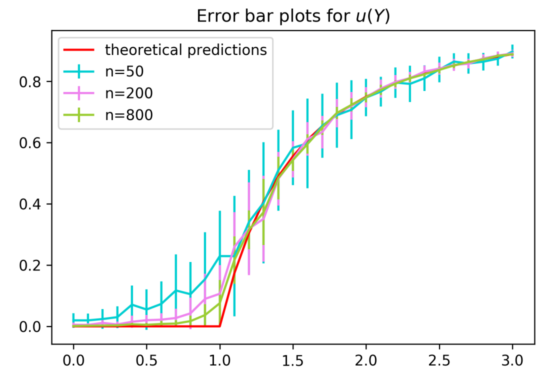
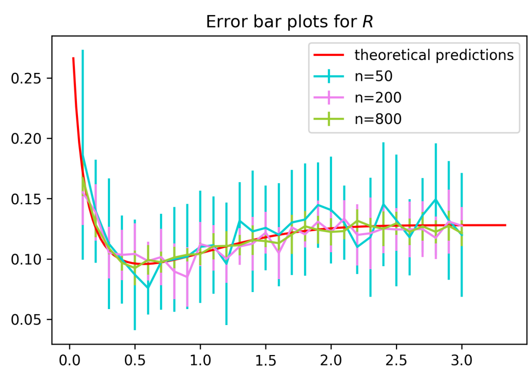
Edited by Taejoo Ahn. Exercise 1-5, 7 is by Taejoo Ahn. Exercise 6 is by Alexander Tsigler.

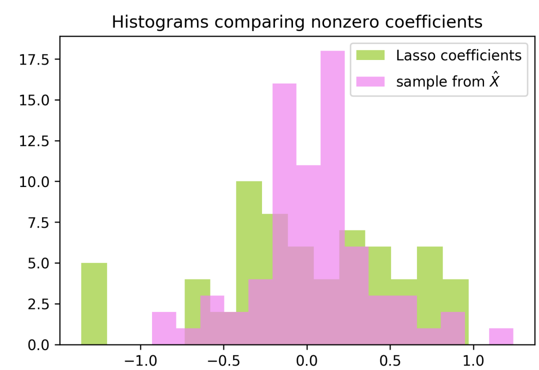
Exercise 1.



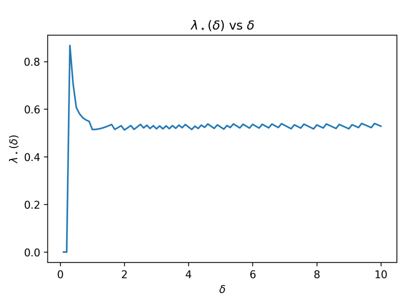


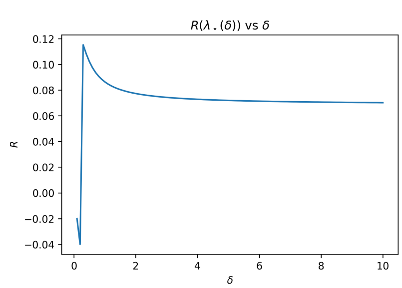
Exercise 2.

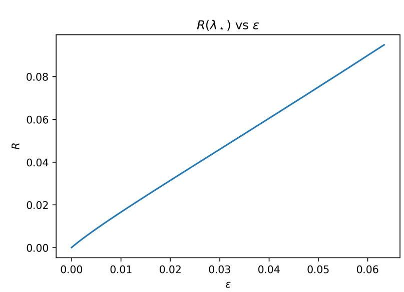




Exercise 3.







Exercise 4.

For $1 \le i \le n$, we have

$$\begin{aligned}
\partial_i \Phi(\boldsymbol{\lambda}) &= \frac{\partial_i \int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)} \\
&= \int_\Omega M_i(\sigma)\left(\frac{\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}\right) \\
&= \int_\Omega M_i(\sigma)\mu_{\boldsymbol{\lambda}}(d\sigma) = \langle M_i\rangle_{\boldsymbol{\lambda}}
\end{aligned}$$

Also for $1 \le i, j \le n$, we have

$$\begin{aligned}
\partial_i\partial_j \Phi(\boldsymbol{\lambda}) &= \partial_i \int_\Omega M_j(\sigma)\left(\frac{\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}\right) \\
&= \frac{\partial_i \int_\Omega M_j(\sigma)\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)} \\
&- \int_\Omega M_j(\sigma)\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)\frac{\partial_i \int_\Omega(\sigma)\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{(\int_\Omega(\sigma)\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma))^2} \\
&= \int_\Omega M_i(\sigma)M_j(\sigma)\left(\frac{\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}\right) \\
&- \int_\Omega M_j(\sigma)\left(\frac{\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}\right)\int_\Omega M_i(\sigma)\left(\frac{\exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}{\int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\sigma)\rangle\}\mu_0(d\sigma)}\right) \\
&= \langle M_i M_j\rangle_{\boldsymbol{\lambda}} - \langle M_i\rangle_{\boldsymbol{\lambda}}\langle M_j\rangle_{\boldsymbol{\lambda}}
\end{aligned}$$

Thus we conclude (1) $\nabla_{\boldsymbol{\lambda}}\Phi(\boldsymbol{\lambda}) = \langle M\rangle_{\boldsymbol{\lambda}}$ and (2) $\nabla^2_{\boldsymbol{\lambda}}\Phi(\boldsymbol{\lambda}) = \langle MM^\top\rangle_{\boldsymbol{\lambda}} - \langle M\rangle_{\boldsymbol{\lambda}}\langle M^\top\rangle_{\boldsymbol{\lambda}}$.

Exercise 5.

STEP 1. For fixed $\lambda > 0$, G_n is Lipschitz in $\boldsymbol{X}$.
Let $\sigma_1, \dots, \sigma_n (\geq 0)$ be eigenvalues of $\boldsymbol{X}\boldsymbol{X}^\top$. Then we have

$$
\begin{aligned}
\|\partial_{\boldsymbol{X}} G_n(\lambda, \boldsymbol{X})\|_F^2 &= \|\frac{2}{n}\boldsymbol{X}^\top(\lambda \boldsymbol{I}_n + \boldsymbol{X}\boldsymbol{X}^\top)^{-1}\|_F^2 \\
&= \frac{4}{n^2} tr\left((\lambda \boldsymbol{I}_n + \boldsymbol{X}\boldsymbol{X}^\top)^{-1}(\boldsymbol{X}\boldsymbol{X}^\top)(\lambda \boldsymbol{I}_n + \boldsymbol{X}\boldsymbol{X}^\top)^{-1}\right) \\
&= \frac{4}{n^2}\sum_{i=1}^n \frac{\sigma_i}{(\lambda + \sigma_i)^2} \\
&\leq \frac{4}{n^2}\sum_{i=1}^n \frac{1}{4\lambda} = \frac{1}{n\lambda}
\end{aligned}
$$

Where the third equality comes from the fact that the trace of a symmetric reel matrix is sum of their eigenvalues.

STEP 2. For any $\epsilon > 0$, there exists $u < \infty$, such that

$$
\lim_{n\to\infty} \mathbb{P}\left(\sup_{\lambda \geq u} |G_n(\lambda) - \log(\lambda)| \geq \epsilon\right) = 0,
$$

and

$$
\lim_{n\to\infty} \sup_{\lambda \geq u} |\mathbb{E}[G_n(\lambda)] - \log(\lambda)| \leq \epsilon.
$$

Let $u = \frac{4}{\epsilon^2}$. Note that $\log(\lambda) = G_n(\lambda, 0)$ so that $|G_n(\lambda) - \log(\lambda)| \leq \sqrt{\frac{1}{n\lambda}}\|\boldsymbol{X}\|_F$. Therefore using Jensen's inequality, we have for any $\lambda \geq u$,

$$
\begin{aligned}
|\mathbb{E}[G_n(\lambda)] - \log(\lambda)| &\leq \mathbb{E}[|G_n(\lambda) - \log(\lambda)|] \\
&\leq \mathbb{E}\left[\sqrt{\frac{1}{n\lambda}}\|\boldsymbol{X}\|_F\right] \\
&\leq \sqrt{\frac{1}{n\lambda}}\sqrt{\mathbb{E}[\|\boldsymbol{X}\|_F^2]} = \sqrt{\frac{1}{\lambda}} < \epsilon
\end{aligned}
$$

Also, with same logic, for fixed $\boldsymbol{X}$, $\{\sup_{\lambda \geq u} |G_n(\lambda) - \log(\lambda)| \geq \epsilon\} \subseteq \{\|\boldsymbol{X}\|_F \geq \sqrt{nu}\epsilon\}$, so that

$$
\begin{aligned}
\mathbb{P}\left(\sup_{\lambda \geq u} |G_n(\lambda) - \log(\lambda)| \geq \epsilon\right) &\leq \mathbb{P}\left(\|\boldsymbol{X}\|_F^2 \geq nu\epsilon^2\right) \\
&= \mathbb{P}(Z \geq 4dn) \\
&\leq e^{-dn}
\end{aligned}
$$

where $Z \sim \chi^2_{dn}$, and the last inequality comes from the concentration inequality of the chi-square distribution. Therefore, taking limit $n \to \infty, n/d \to \gamma$, we have both desired result.

STEP 3. For any $\lambda_0 > 0$ and $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\lambda_0\le\lambda} |G_n(\lambda, \boldsymbol{X}) - \mathbb{E}[G_n(\lambda, \boldsymbol{X})]| \ge \epsilon\right) = 0$$

Combined with STEP 2, it is enough to show that

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\lambda_0\le\lambda\le u} |G_n(\lambda, \boldsymbol{X}) - \mathbb{E}[G_n(\lambda, \boldsymbol{X})]| \ge \epsilon\right) = 0$$

First, for any fixed λ, using $G_n(\lambda, \boldsymbol{X})$ is Lipchitz, we have:

$$\mathbb{P}(|G_n(\lambda, \boldsymbol{X}) - \mathbb{E}G_n(\lambda, \boldsymbol{X})| \ge t) \le 2e^{-\lambda ndt^2/2}$$

(Let $f(\tilde{\boldsymbol{X}}) := G_n(\lambda, \tilde{\boldsymbol{X}}/\sqrt{d})$, which is $1/\sqrt{\lambda nd}$-Lipschitz. Then let $\tilde{\boldsymbol{X}} := \sqrt{d}\boldsymbol{X}$ so that $\tilde{\boldsymbol{X}}$ has i.i.d standard Gaussian entries. Then we can apply Gaussian concentration inequality on $f(\tilde{\boldsymbol{X}}) \overset{d}{\equiv} G_n(\lambda, \boldsymbol{X})$ to get the concentration bound above.) Now let $\mathcal{G}(\epsilon)$ be a ϵ cover of $\mathcal{G} := \{G(\lambda, \cdot)|\lambda \in [\lambda_0, u]\}$, where $N(\epsilon) = |\mathcal{G}(\epsilon)|$ is the covering number of $\mathcal{G}$, where we use ∞-norm as the metric. Now for each $g \in \mathcal{G}$, let $\tilde{g} \in \mathcal{G}(\epsilon)$ be an element in g such that $\|g - \tilde{g}\| \le \epsilon$. Then we have

$$\begin{aligned}
\sup_{\lambda\in[\lambda_0,u]} |G_n(\lambda, \boldsymbol{X}) - \mathbb{E}[G_n(\lambda, \boldsymbol{X})]| &= \sup_{g\in\mathcal{G}} |g(\boldsymbol{X}) - \mathbb{E}[g(\boldsymbol{X})]| \\
&\le \sup_{g\in\mathcal{G}} (|g(\boldsymbol{X}) - \tilde{g}(\boldsymbol{X})| + |\tilde{g}(\boldsymbol{X}) - \mathbb{E}[\tilde{g}(\boldsymbol{X})]| + |\mathbb{E}[\tilde{g}(\boldsymbol{X})] - \mathbb{E}[g(\boldsymbol{X})]|) \\
&\le 2\epsilon + \sup_{g\in\mathcal{G}(\epsilon)} |g(\boldsymbol{X}) - \mathbb{E}[g(\boldsymbol{X})]|
\end{aligned}$$

Therefore, using union bound, for any $\epsilon \in (0, t/2)$, we have

$$\begin{aligned}
\mathbb{P}(\sup_{\lambda\in[\lambda_0,u]} |G_n(\lambda, \boldsymbol{X}) - \mathbb{E}[G_n(\lambda, \boldsymbol{X})]| \ge t) &\le \mathbb{P}(\sup_{g\in\mathcal{G}(\epsilon)} |g(\boldsymbol{X}) - \mathbb{E}[g(\boldsymbol{X})]| \ge t - 2\epsilon) \\
&\le 2N(\epsilon)e^{-\lambda nd(t-2\epsilon)^2/2}
\end{aligned}$$

Note that $|\partial_\lambda G_n(\lambda, \boldsymbol{X})| = \frac{1}{n} tr((\lambda \boldsymbol{I}_n + \boldsymbol{X}\boldsymbol{X}^\top)^{-1}) \le \frac{1}{\lambda_0}$, and thus $N(\epsilon) \le \frac{u-\lambda_0}{\lambda_0\epsilon}$. Thus taking $\epsilon = t/4$ gives

$$\mathbb{P}(\sup_{\lambda\in[\lambda_0,u]} |G_n(\lambda, \boldsymbol{X}) - \mathbb{E}[G_n(\lambda, \boldsymbol{X})]| \ge t) \le \frac{8(u-\lambda_0)}{\lambda_0 t} e^{-\lambda ndt^2/8}$$

and thus taking limit $n \to \infty, n/d \to \gamma$ gives the desired result.

Exercise 6.

Question 1

$$\begin{aligned}
F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) =& \frac{1}{n\beta}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})\\
=& \frac{1}{n\beta}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma}),
\end{aligned}$$

where $\boldsymbol{X}$ is a matrix with i.i.d. $\mathcal{N}(0,1/2n)$ entries.

First, let's compute the gradient of $F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})$ in $\boldsymbol{X}$:

$$\begin{aligned}
\nabla_{\boldsymbol{X}}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) =& \frac{1}{n\beta}\nabla_{\boldsymbol{X}}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})\\
=& \frac{1}{n\beta}\frac{\int_{\mathbb{S}^{n-1}}\nabla_{\boldsymbol{X}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}\\
=& \frac{\int_{\mathbb{S}^{n-1}}\boldsymbol{\sigma}\boldsymbol{\sigma}^\top\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}
\end{aligned}$$

— note that this is expectation of $\boldsymbol{\sigma}\boldsymbol{\sigma}^\top$ with respect to the measure on the sphere whose density is proportional to $\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})$. Since $\|\boldsymbol{\sigma}\boldsymbol{\sigma}^\top\|_F=1$ for any $\boldsymbol{\sigma}\in\mathbb{S}^{n-1}$ and since expectation of a norm is always greater or equal than the norm of expectation, we obtain

$$\|\nabla_{\boldsymbol{X}}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})\|_F\le 1$$

Thus, by Gaussian Lipschitz concentration

$$\mathbb{P}(|F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})-\mathbb{E}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})|>t)\le 2\exp(-nt^2).$$

Now, as in the previous exercise, we also compute derivatives in λ and β to obtain uniform results:

$$\begin{aligned}
\frac{\partial}{\partial\lambda}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) =& \frac{\partial}{\partial\lambda}\frac{1}{n\beta}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})\\
=& \frac{\int_{\mathbb{S}^{n-1}}\langle\boldsymbol{\sigma},\boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{\sigma}\rangle/2\cdot\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}\\
=& \langle(\boldsymbol{v}^\top\boldsymbol{\sigma})^2\rangle_{\beta,\lambda,n}/2,
\end{aligned}$$

where $\langle\cdot\rangle_{\beta,\lambda,n}$ denote average w.r.t. the probability measure on $\mathbb{S}^{n-1}$ proportional to $\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})$.

Since $\boldsymbol{v}$ and $\boldsymbol{\sigma}$ are on the sphere, we see that

$$\left|\frac{\partial}{\partial\lambda}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})\right|\le 1/2.$$

Now we take derivative in β:

$$\begin{aligned}\frac{\partial}{\partial\beta}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) =&\frac{\partial}{\partial\beta}\frac{1}{n\beta}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})\\ =&-\frac{1}{n\beta^2}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})\\ &+\frac{1}{\beta}\frac{\int_{\mathbb{S}^{n-1}}\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\cdot\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}\end{aligned}$$

Note that since β and λ belongs to compact regios $\mathcal{B}$ and Λ separated from zero, $\boldsymbol{\sigma}$ and $\boldsymbol{v}$ have unit norms, and $\boldsymbol{W}$ has norm bounded by a constant with high probability (probability goes to 1 as n goes to infinity), the following inequalities imply that $F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})$ is Lipschitz in β with high probability:

$$\begin{aligned}&\frac{1}{n\beta^2}\log\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})\\ \le&\frac{1}{n\beta^2}\log\max_{\boldsymbol{\sigma}\in\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\\ =&\frac{1}{n\beta^2}\max_{\boldsymbol{\sigma}\in\mathbb{S}^{n-1}}\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\\ \le&\frac{1}{\min_{\mathcal{B}}\beta}(\max_{\Lambda}\lambda+\|\boldsymbol{W}\|)/2\end{aligned}$$

— bounded by a constant with probability that goes to one as n goes to infinity (the constant only depends on Λ and $\mathcal{B}$, but not on n).

$$\begin{aligned}&\frac{1}{\beta}\frac{\int_{\mathbb{S}^{n-1}}\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\cdot\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}}\exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}\\ \le&\frac{1}{\min_{\mathcal{B}}\beta}\max_{\boldsymbol{\sigma}\in\mathbb{S}^{n-1}}\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\top+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\\ \le&\frac{1}{\min_{\mathcal{B}}\beta}(\max_{\Lambda}\lambda+\|\boldsymbol{W}\|)/2\end{aligned}$$

— bounded by a constant with probability that goes to one as n goes to infinity (the constant only depends on Λ and $\mathcal{B}$, but not on n).

The last thing we need is to bound the Lipschitz constant of the expectation of $F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})$ w.r.t. variables β and λ. $F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})$ is a.s. 1-Lipschitz in λ, so it's expectation is also 1-Lipschitz. When it comes to dependence on β, we derived that

$$|F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})-F_n(\beta',\lambda,\boldsymbol{v},\boldsymbol{W})|\le\frac{1}{\min_{\mathcal{B}}\beta}(\max_{\Lambda}\lambda+\|\boldsymbol{W}\|)|\beta-\beta'|.$$

Taking expectation we get

$$|\mathbb{E}F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})-\mathbb{E}F_n(\beta',\lambda,\boldsymbol{v},\boldsymbol{W})|\le\mathbb{E}|F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})-F_n(\beta',\lambda,\boldsymbol{v},\boldsymbol{W})|\le\frac{1}{\min_{\mathcal{B}}\beta}(\max_{\Lambda}\lambda+\mathbb{E}\|\boldsymbol{W}\|)|\beta-$$

Since $\mathbb{E}\|\boldsymbol{W}\|$ is bounded by a constant independent of n, we see that $\mathbb{E}F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})$ is also Lipschitz with constant Lipcshitz constant.

The rest of the argument is analogous to the previous exercise: for any fixed compacts Λ and $\mathcal{B}$ that are separated from zero and fixed ϵ do the following:

1. Fix $\delta > 0$.
2. For n large enough with probability at least $1 - \delta/2$ $F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})$ and $\mathbb{E}F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})$ have bounded (by a constant which doesn't depend on n) Lipschitz constants, so the supremum of $F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) - \mathbb{E}F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})$ over $\mathcal{B} \times \Lambda$ doesn't exceed $\epsilon/2$ plus supremum over a ϵ-net over $\mathcal{B} \times \Lambda$.
3. Since the ϵ-net is a finite set and we have concentration at each point, for n large enough with probability at least $1 - \delta/2$ supremum over ϵ-net doesn't exceed $\epsilon/2$.
4. Thus, for any δ if n is large enough then
$$\mathbb{P}(\sup_{\mathcal{B}\times\Lambda} |F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) - \mathbb{E}F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})| > \epsilon) \leq 1 - \delta.$$
5. Since we can take δ arbitrary small, we obtain
$$\lim_{n\to\infty} \mathbb{P}(\sup_{\mathcal{B}\times\Lambda} |F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) - \mathbb{E}F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})| > \epsilon) = 0.$$

This finishes the proof.

Question 2

We already showed in the previous part that

$$\frac{\partial}{\partial\lambda} F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) = \langle(\boldsymbol{v}^\top\boldsymbol{\sigma})^2\rangle_{\beta,\lambda,n}/2,$$

where $\langle\cdot\rangle_{\beta,\lambda,n}$ denote average w.r.t. the probability measure on $\mathbb{S}^{n-1}$ proportional to $\exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})$. If we plug $\beta = \lambda$ the measure becomes proportional to $\exp\{n\lambda\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})$ — exactly the posterior distribution over $\boldsymbol{\sigma}$. Thus, we have

$$\frac{\partial}{\partial\lambda} F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})|_{\beta=\lambda} = \boldsymbol{v}^\top\langle\boldsymbol{\sigma}\boldsymbol{\sigma}^\top\rangle_{p(d\boldsymbol{\sigma}|\boldsymbol{Y})}\boldsymbol{v}/2 = \boldsymbol{v}^\top\hat{\boldsymbol{V}}(\boldsymbol{Y})\boldsymbol{v}/2.$$

Now let's compute the second derivative:

$$\begin{aligned}
\frac{\partial^2}{\partial\lambda^2} F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) =& \frac{\partial}{\partial\lambda} \frac{\int_{\mathbb{S}^{n-1}} \langle\boldsymbol{\sigma}, \boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{\sigma}\rangle/2 \cdot \exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}} \exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})} \\
=& \frac{n\beta/4 \int_{\mathbb{S}^{n-1}} \langle\boldsymbol{\sigma}, \boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{\sigma}\rangle^2 \cdot \exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}} \exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})} \\
&- n\beta/4 \left(\frac{\int_{\mathbb{S}^{n-1}} \langle\boldsymbol{\sigma}, \boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{\sigma}\rangle \cdot \exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}{\int_{\mathbb{S}^{n-1}} \exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})}\right)^2
\end{aligned}$$

We see that$(\partial^2/\partial\lambda^2)F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})/(n\beta)$ is the variance of quantity $\boldsymbol{v}^\top\boldsymbol{\sigma})^2$ w.r.t the probability measure on $\mathbb{S}^{n-1}$ which is proportional to $\exp\{n\beta\langle\boldsymbol{\sigma}, (\lambda\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(d\boldsymbol{\sigma})$. Since variance is always non-negative, $F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})$ is convex in λ.

Question 3

Suppose f is a convex function of $[a, b]$ and $\sup_{[a,b]} |f(x)| \leq \epsilon$. Let's take some point $c \in (a + \delta, b - \delta)$. Since f is convex, $f(a) \geq f(c) + f'(c)(a - c)$ and $f(b) \geq f(c) + f'(c)(b - c)$. Thus, we can write

$$-2\epsilon/\delta \leq \frac{f(c) - f(a)}{c - a} \leq f'(c) \leq \frac{f(b) - f(c)}{b - c} \leq 2\epsilon/\delta.$$

This implies that uniform convergence of convex functions f_n to zero on $[a, b]$ implies uniform convergence of f'_n on $[a + \delta, b - \delta]$.

Now let's take Λ' to be a compact set that doesn't contain zero, but contains δ-vicinity of Λ for some $\delta > 0$. From the previous questions we know that

$$\lim_{n\to\infty} \mathbb{P}(\sup_{(\beta,\lambda)\in\Lambda'\times\Lambda'} |F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) - \mathbb{E}F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})| > \epsilon) = 0.$$

From the statement above and the fact that $F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})$ is convex in λ, we obtain

$$\lim_{n\to\infty} \mathbb{P}(\sup_{(\beta,\lambda)\in\Lambda\times\Lambda} |\partial_\lambda F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W}) - \mathbb{E}\partial_\lambda F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})| > 2\epsilon/\delta) = 0,$$

(note that we swapped expectation and differentiation without a proof).

And thus

$$\lim_{n\to\infty} \mathbb{P}(\sup_{(\beta,\lambda)\in\Lambda\times\Lambda} |\partial_\lambda F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})_{\beta=\lambda} - \mathbb{E}\partial_\lambda F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})_{\beta=\lambda}| > 2\epsilon/\delta) = 0,$$

Since ϵ is arbitrary and $\partial_\lambda F_n(\beta, \lambda, \boldsymbol{v}, \boldsymbol{W})|_{\beta=\lambda} = \boldsymbol{v}^\top \hat{\boldsymbol{V}}(\boldsymbol{Y})\boldsymbol{v}/2$ the desired statement follows.

Exercise 7.

Question 1. Let $H_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}) + \langle \boldsymbol{\sigma}, \boldsymbol{\lambda} \rangle$, then

$$\begin{aligned} \nabla_{\boldsymbol{\lambda}_i} F(\beta, \boldsymbol{\lambda}) &= \sum_{\boldsymbol{\sigma} \in \Omega_d} \left(-\frac{1}{\beta}\right)(-\beta\sigma_i) \frac{\exp\{-\beta H_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})\}}{\sum_{\boldsymbol{\sigma}\in\Omega_d} \exp\{-\beta H_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})\}} \\ &= \sum_{\boldsymbol{\sigma}\in\Omega_d} \sigma_i \mu_{\beta,\boldsymbol{\lambda}}(\boldsymbol{\sigma}) = \langle \sigma_i \rangle_{\beta,\boldsymbol{\lambda}} \end{aligned}$$

and thus $\nabla_{\boldsymbol{\lambda}} F(\beta, \boldsymbol{\lambda})|_{\boldsymbol{\lambda}=0} = \langle \boldsymbol{\sigma} \rangle_\beta$

Question 2. Let $H_\lambda(\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}) + \lambda \|\boldsymbol{\sigma}\|_2^2$, then

$$\begin{aligned} \nabla_\lambda F(\beta, \lambda) &= \left(-\frac{1}{\beta}\right)(-\beta\|\boldsymbol{\sigma}\|_2^2) \frac{\exp\{-\beta H_\lambda(\boldsymbol{\sigma})\}}{\sum_{\boldsymbol{\sigma}\in\Omega_d} \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\}} \\ &= \sum_{\boldsymbol{\sigma}\in\Omega_d} \|\boldsymbol{\sigma}\|_2^2 \mu_{\beta,\lambda}(\boldsymbol{\sigma}) = \langle \|\boldsymbol{\sigma}\|_2^2 \rangle_{\beta,\lambda} \end{aligned}$$

and thus $\nabla_\lambda F(\beta, \lambda)|_{\lambda=0} = \langle \|\boldsymbol{\sigma}\|_2^2 \rangle_\beta$

Question 3. Let $H_\lambda(\boldsymbol{\sigma}') = H(\boldsymbol{\sigma}_1) + H(\boldsymbol{\sigma}_2) + \lambda\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle$, then

$$\begin{aligned} \nabla_\lambda F(\beta, \lambda)|_{\lambda=0} &= \sum_{\boldsymbol{\sigma}'\in\Omega'} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle \frac{\exp\{-\beta H(\boldsymbol{\sigma}_1) - \beta H(\boldsymbol{\sigma}_2)\}}{\sum_{\boldsymbol{\sigma}'\in\Omega'} \exp\{-\beta H(\boldsymbol{\sigma}_1) - \beta H(\boldsymbol{\sigma}_2)\}} \\ &= \sum_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\Omega} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle \frac{\exp\{-\beta H(\boldsymbol{\sigma}_1)}{\sum_{\boldsymbol{\sigma}_1\in\Omega_d} \exp\{-\beta H(\boldsymbol{\sigma}_1)\}} \frac{\exp\{-\beta H(\boldsymbol{\sigma}_2)}{\sum_{\boldsymbol{\sigma}_2\in\Omega_d} \exp\{-\beta H(\boldsymbol{\sigma}_2)\}} \\ &= \sum_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\Omega} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle \mu_\beta(\boldsymbol{\sigma}_1)\mu_\beta(\boldsymbol{\sigma}_2) \\ &= \sum_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\Omega} \sum_{i=1}^n \sigma_{1i}\sigma_{2i} \mu_\beta(\boldsymbol{\sigma}_1)\mu_\beta(\boldsymbol{\sigma}_2) \\ &= \sum_{i=1}^n \sum_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\Omega} \sigma_{1i}\sigma_{2i} \mu_\beta(\boldsymbol{\sigma}_1)\mu_\beta(\boldsymbol{\sigma}_2) \\ &= \sum_{i=1}^n \langle \boldsymbol{\sigma}_i \rangle_\beta \cdot \langle \boldsymbol{\sigma}_i \rangle_\beta = \|\langle \boldsymbol{\sigma} \rangle_\beta\|_2^2 \end{aligned}$$

Question 4. Let $\Omega' = \{\pm 1\}^{d\times p}$, and $H_\lambda(\boldsymbol{\sigma}') = \sum_{i=1}^p H(\boldsymbol{\sigma}_i) + \lambda \prod_{i=1}^p \langle \boldsymbol{A}_i, \boldsymbol{\sigma}_i \rangle$ then with same logic with Question 3, we have

$$
\begin{aligned}
\nabla_\lambda F(\beta,\lambda)|_{\lambda=0} &= \sum_{\boldsymbol{\sigma}'\in\Omega'}\prod_{i=1}^{p}\langle \boldsymbol{A}_i, \boldsymbol{\sigma}_i\rangle \frac{\prod_i \exp\{-\beta H(\boldsymbol{\sigma}_i)\}}{\prod_i \sum_{\boldsymbol{\sigma}_i\in\Omega}\exp\{-\beta H(\boldsymbol{\sigma}_i)\}} \\
&= \sum_{\boldsymbol{\sigma}'\in\Omega'}\prod_{i=1}^{p}\langle \boldsymbol{A}_i, \boldsymbol{\sigma}_i\rangle \prod_i \exp\{-\beta H(\boldsymbol{\sigma}_i)\}/Z(\beta) \\
&= \sum_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_p\in\Omega}\prod_{i=1}^{p}\left(\langle \boldsymbol{A}_i, \boldsymbol{\sigma}_i\rangle \mu_\beta(\boldsymbol{\sigma}_i)\right) \\
&= \prod_{i=1}^{p}\langle\langle \boldsymbol{A}_i, \boldsymbol{\sigma}\rangle\rangle_\beta \\
&= \prod_{i=1}^{p}\langle \boldsymbol{A}_i, \langle\boldsymbol{\sigma}\rangle_\beta\rangle = \langle \boldsymbol{A}, \langle\boldsymbol{\sigma}\rangle_\beta^{\otimes p}\rangle
\end{aligned}
$$