Electronic Journal of Statistics
Vol. 17 (2023) 19–69
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2098

Convergence properties of data augmentation algorithms for high-dimensional robit regression

Sourav Mukherjee[1], **Kshitij Khare**[1] **and Saptarshi Chakraborty**[2]

[1] *Department of Statistics, University of Florida*
e-mail: souravmukherjee@ufl.edu

[2] *Department of Biostatistics, State University of New York at Buffalo*
e-mail: kdkhare@stat.ufl.edu; chakrab2@buffalo.edu

Abstract: The logistic and probit link functions are the most common choices for regression models with a binary response. However, these choices are not robust to the presence of outliers/unexpected observations. The robit link function, which is equal to the inverse CDF of the Student's t-distribution, provides a robust alternative to the probit and logistic link functions. A multivariate normal prior for the regression coefficients is the standard choice for Bayesian inference in robit regression models. The resulting posterior density is intractable and a Data Augmentation (DA) Markov chain is used to generate approximate samples from the desired posterior distribution. Establishing geometric ergodicity for this DA Markov chain is important as it provides theoretical guarantees for asymptotic validity of MCMC standard errors for desired posterior expectations/quantiles. Previous work [16] established geometric ergodicity of this robit DA Markov chain assuming (i) the sample size n dominates the number of predictors p, and (ii) an additional constraint which requires the sample size to be bounded above by a fixed constant which depends on the design matrix X. In particular, modern high-dimensional settings where $n < p$ are not considered. In this work, we show that the robit DA Markov chain is trace-class (i.e., the eigenvalues of the corresponding Markov operator are summable) for arbitrary choices of the sample size n, the number of predictors p, the design matrix X, and the prior mean and variance parameters. The trace-class property implies geometric ergodicity. Moreover, this property allows us to conclude that the sandwich robit chain (obtained by inserting an inexpensive extra step in between the two steps of the DA chain) is strictly better than the robit DA chain in an appropriate sense, and enables the use of recent methods to estimate the spectral gap of trace class DA Markov chains.

MSC2020 subject classifications: Primary 60J05, 60J20.
Keywords and phrases: Markov chain Monte Carlo, geometric ergodicity, high-dimensional binary regression, robust model, trace class.

Received December 2021.

Contents

arXiv: 2112.10349

1. Introduction

Consider a regression setting with n independent binary responses $Y_1, Y_2, \cdots, Y_n$ and corresponding predictor vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \in \mathbb{R}^p$, such that

$$P(Y_i = 1 \mid \boldsymbol{\beta}) = F(\mathbf{x}_i^T \boldsymbol{\beta}),$$

for $1 \leq i \leq n$. Here F is a strictly increasing cumulative distribution function, and $G = F^{-1}$ is referred to as the link function. Two popular choices of F are given by $F(x) = \frac{e^x}{1+e^x}$ (logistic link) and $F(x) = \Phi(x)$ (probit link) where $\Phi(x)$ denotes the standard normal CDF. It is well known that estimates of $\boldsymbol{\beta}$ produced for both these choices are not robust to outliers [13, 5]. To address such settings, F is set to be the CDF of the Student's t-distribution, and the corresponding model is referred to as the robit regression model [9]. For binary responses, an outlier is an unexpected observation with large value(s) of the predictor(s) and a misclassified response, and the robit model effectively down-weights such outliers to produce a better fit [6].

Following [16, 1] we consider a Bayesian robit regression model specified as follows.

$$\begin{aligned}
&P(Y_i = 1 \mid \boldsymbol{\beta}) = F_\nu(\mathbf{x}_i^T \boldsymbol{\beta}) \text{ for } 1 \leq i \leq n, \\
&\boldsymbol{\beta} \sim \mathcal{N}_p\left(\boldsymbol{\beta}_a, \Sigma_a^{-1}\right),
\end{aligned}$$

where F_ν denotes the CDF of the Student's t-distribution with ν degrees of freedom (with location 0, scale 1) and $\mathcal{N}_p$ denotes the p variate normal distribution. Let $\mathbf{Y}$ denote the response vector, and let $\pi(\boldsymbol{\beta} \mid \mathbf{y})$ denote the posterior density of $\boldsymbol{\beta}$ given $\mathbf{Y} = \mathbf{y}$. As demonstrated in [16], the posterior density $\pi(\boldsymbol{\beta} \mid \mathbf{y})$ is intractable in the sense that relevant posterior expectations are ratios of two intractable integrals and are not available in closed form. Also, generating IID samples from this density is computationally infeasible even for moderately large values of p. To resolve this, [16] develops a clever and effective Data Augmentation (DA) approach which can be used to construct a computationally tractable Markov chain which has $\pi(\boldsymbol{\beta} \mid \mathbf{y})$ as its stationary density. We describe this Markov chain below.

Let $t_\nu(\mu, \sigma)$ denote the Student's t-distribution with ν degrees of freedom, location μ and scale σ. Consider unobserved latent variables

$$(Z_1, \lambda_1), (Z_2, \lambda_2), \cdots, (Z_n, \lambda_n)$$

which are mutually independent and satisfy $Z_i \mid \lambda_i \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, 1/\lambda_i)$ and $\lambda_i \sim \text{Gamma}(\nu/2, \nu/2)$. Then, it can be shown that the marginal distribution of Z_i is given by $Z_i \sim t_\nu(\mathbf{x}_i^T \boldsymbol{\beta}, 1)$. If Y_i is defined as the indicator of Z_i taking positive values, i.e., $Y_i = 1_{\{Z_i > 0\}}$, then $P(Y_i = 1 \mid \boldsymbol{\beta}) = P(z_i > 0) = F_\nu(\mathbf{x}_i^T \boldsymbol{\beta})$, which is consistent with the robit regression model specified in (1.1). Straightforward calculations (see [16]) now show the following.

- $(Z_1, \lambda_1), (Z_2, \lambda_2), \cdots, (Z_n, \lambda_n)$ are conditionally independent given $\boldsymbol{\beta}, \mathbf{Y} = \mathbf{y}$. Also,

$$\begin{aligned} &Z_i \mid \boldsymbol{\beta}, \mathbf{y} \sim Tt_\nu(\mathbf{x}_i^T \boldsymbol{\beta}, y_i), \\ &\lambda_i \mid Z_i = z_i, \boldsymbol{\beta}, \mathbf{y} \sim \text{Gamma}\left(\frac{\nu+1}{2}, \frac{\nu + (z_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{2}\right). \end{aligned}$$

  Here $Tt_v(\mathbf{x}_i^T \boldsymbol{\beta}, y_i)$ denotes the t_ν distribution with location $\mathbf{x}_i^T \boldsymbol{\beta}$ and scale 1, truncated to $\mathbb{R}_+$ if $y_i = 1$ and to $\mathbb{R}_-$ if $y_i = 0$.
- Let $\mathbf{Z} = (Z_1, \cdots, Z_n)^T$, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)^T$ and Λ denote a diagonal matrix whose diagonal is given by entries of $\boldsymbol{\lambda}$. The conditional distribution of $\boldsymbol{\beta}$ given $\mathbf{Z} = \mathbf{z}, \boldsymbol{\lambda}, \mathbf{Y} = \mathbf{y}$ is

$$\mathcal{N}_p\left((X^T \Lambda X + \Sigma_a)^{-1}(X^T \Lambda \mathbf{z} + \Sigma_a \boldsymbol{\beta}_a),\ (X^T \Lambda X + \Sigma_a)^{-1}\right).$$

These observations are used in [16] to construct a DA Markov chain $\{\boldsymbol{\beta}^{(m)}\}_{m \geq 0}$ on $\mathbb{R}^p$ (with stationary density $\pi(\boldsymbol{\beta} \mid \mathbf{y})$) whose one step transition from $\boldsymbol{\beta}^{(m)}$ to $\boldsymbol{\beta}^{(m+1)}$ is described in the following Algorithm 1. Harris ergodicity of the

Algorithm 1: $(m+1)$-st Iteration of the Robit Data Augmentation Algorithm

1. Make independent draws from the $Tt_\nu(\mathbf{x}_i^T \boldsymbol{\beta}^{(m)}, y_i)$ distributions for $1 \leq i \leq n$. Denote the respective draws by $z_1, z_2, \cdots, z_n$. Draw λ_i from the $\text{Gamma}\left(\frac{\nu+1}{2}, \frac{\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}^{(m)}\right)^2}{2}\right)$ distribution.
2. Draw $\boldsymbol{\beta}^{(m+1)}$ from the $\mathcal{N}_p\left((X^T \Lambda X + \Sigma_a)^{-1}(X^T \Lambda \mathbf{z} + \Sigma_a \boldsymbol{\beta}_a),\ (X^T \Lambda X + \Sigma_a)^{-1}\right)$ distribution.

robit DA Markov chain $\{\boldsymbol{\beta}^{(m)}\}_{m \geq 0}$ obtained through Algorithm 1 is established in [16]. Suppose a posterior expectation $E_{\pi(\cdot|\mathbf{y})}[h(\boldsymbol{\beta})]$ (assumed to exist) is of interest. Then by Harris ergodicity, the cumulative averages $\frac{1}{m+1} \sum_{r=0}^m h(\boldsymbol{\beta}^{(r)})$ converge to $E_{\pi(\cdot|\mathbf{y})}[h(\boldsymbol{\beta})]$ as $m \to \infty$, and can be used to approximate the desired posterior expectation. However, any such approximation is useful only with an estimate of the associated error. The standard approach for obtaining such error estimates is to establish a Markov chain central limit theorem (CLT) which guarantees that

$$\sqrt{m}\left(\frac{1}{m+1}\sum_{r=0}^{m} h(\boldsymbol{\beta}^{(r)}) - E_{\pi(\cdot|\mathbf{y})}[h(\boldsymbol{\beta})]\right) \xrightarrow{D} \mathcal{N}(0, \sigma_h^2)$$

as $m \to \infty$, and then construct a consistent estimate $\hat{\sigma}_h$ of the asymptotic standard deviation σ_h. A key sufficient condition for establishing a Markov chain CLT is *geometric ergodicity*. A Markov chain is geometrically ergodic if the total variation distance between its distribution after m steps and the stationary distribution converges to 0 as $m \to \infty$. To summarize, establishing geometric ergodicity of the robit DA chain is critical for obtaining asymptotically valid standard errors for Markov chain based estimates of posterior quantities.

With this motivation [16] investigated and established geometric ergodicity of the robit DA chain. However, it is assumed that the design matrix is full rank (which implies $n \geq p$ and rules out high-dimensional settings), $\Sigma_a = g^{-1}X^TX$ and that

$$n \leq \frac{g^{-1}\nu}{(\nu+1)(1+2\sqrt{\boldsymbol{\beta}_a^T X^T X \boldsymbol{\beta}_a})}.$$

The last upper bound on n involving the design matrix, the prior mean and covariance and the degrees of freedom ν is in particular very restrictive. While we found that these conditions can be relaxed to some extent by a tighter drift and minorization analysis, the resulting constraints still remain quite restrictive. Geometric ergodicity results for the related probit DA chain (see [2]) do not require such assumptions, and quoting from [16, Page 2469] "Ideally, we would like to be able to say that the DA algorithm is geometrically ergodic for any $n, \nu, \mathbf{y}, X, \boldsymbol{\beta}_a, \Sigma_a$".

In the probit DA setting (see [1]) the latent variables Z_i have a normal distribution, and there is no need to introduce the additional latent variables λ_i. This additional layer of latent variables creates additional complexity in the structure of the robit DA chain which makes the convergence analysis significantly more challenging compared to the probit DA chain analyses undertaken in [17, 2].

It is not clear if the restrictive conditions listed above for geometric ergodicity of the robit DA chain are really necessary or if they are an artifact of the standard drift and minorization technique used in [16] for establishing geometric ergodicity. We take a completely different approach, and focus on investigating the *trace class* property for the robit DA chain. A Markov chain with stationary density π is trace class if the corresponding Markov operator on $L^2(\pi)$ has a countable spectrum and the corresponding eigenvalues are summable. The trace class property implies geometric ergodicity, and can be established by showing that an appropriate integral involving the transition density of the Markov chain is finite (see Section 2). As the main technical contribution of this paper, *we establish that the robit DA chain is trace class for any* $n, p, \nu > 0, \mathbf{y}, X, \boldsymbol{\beta}_a$ *and positive definite* Σ_a. This in particular *establishes geometric ergodicity for any* $n, p, \nu > 0, \mathbf{y}, X, \boldsymbol{\beta}_a$ *and positive definite* Σ_a*, and significantly generalizes existing results* in [16].

The trace class property is much stronger than geometric ergodicity, and the bounding of the relevant integral can get quite involved and challenging (see for example the analysis is [2] and in Section 2). It is therefore not surprising that the conditions needed for establishing geometric ergodicity through the trace class approach have typically been stronger than those needed to establish geometric ergodicity using the drift and minorization approach (for chains where both such analyses have been successful). For example, [2] considers convergence analysis of the probit DA chain with a proper prior for $\boldsymbol{\beta}$ as in (1.1). Using drift and minorization geometric ergodicity is established for all $n, p, \mathbf{y}, X, \boldsymbol{\beta}_a, \Sigma_a$, but the trace class property was only established under some constraints on X and Σ_a (see [2, Theorem 2]). Hence, it is quite interesting that for the robit DA chain the reverse phenomenon holds: the drift and minorization approach, with the drift function chosen in [16], needs stronger conditions to succeed than the trace class approach. Essentially, the additional layer of latent variables λ_i introduced in the robit setting severely hampers the drift and minorization analysis, but with careful additional analysis and direct utilization of the structure of a t-cdf embodying the inverse robit link function, eases the path for showing the finiteness of the relevant trace class integral.

Establishing the trace class property of the DA chain gives additional benefits on top of geometric ergodicity. Using results from [8], it can now be concluded that the sandwich robit DA chain constructed in [16] is also trace class and is strictly better than the robit DA chain (in the sense that the spectrum of the latter strictly dominates the spectrum of the former). Also, the trace class property is a key sufficient condition for using recent approaches in [3, 14] to estimate the spectral gap of Markov chains. The remainder of the paper is organized as follows. Section 2 contains the proof of the trace class property for the robit DA chain. Section 3 provides numerical illustrations of various chains on two real datasets, one with $n \geq p$, and one with $n < p$. Additional mathematical results needed for the proof of the trace class property are provided in an appendix.

2. Trace-class property for the DA chain

Recall from [16] that the DA Markov chain has associated transition density given by

$$\begin{aligned} k\left(\boldsymbol{\beta}, \boldsymbol{\beta}'\right) &= \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^n} \pi\left(\boldsymbol{\beta}' \middle| \boldsymbol{\lambda}, \mathbf{z}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda}, \mathbf{z} \middle| \boldsymbol{\beta}, \mathbf{y}\right) d\mathbf{z}\; d\boldsymbol{\lambda} \\ &= \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^n} \pi\left(\boldsymbol{\beta}' \middle| \boldsymbol{\lambda}, \mathbf{z}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \middle| \mathbf{z}, \boldsymbol{\beta}, \mathbf{y}\right) \pi\left(\mathbf{z} \middle| \boldsymbol{\beta}, \mathbf{y}\right) d\mathbf{z}\; d\boldsymbol{\lambda}. \end{aligned} \tag{2.1}$$

Let $L_0^2(\pi(.|\mathbf{y}))$ denote the space of square-integrable functions with mean zero (with respect to the posterior density $\pi(\boldsymbol{\beta}|\mathbf{y})$). Let K denote the Markov operator on $L_0^2(\pi(.|\mathbf{y}))$ associated with the transition density k. Note that the Markov chain corresponding to k tracks the movement of one component of a two-component DA algorithm. It follows from the results in [10] that the Markov transition density k is reversible with respect to its invariant distribution, and K is a positive, self-adjoint operator. The operator K is *trace class* (see Jörgens [7]) if

$$I := \int_{\mathbb{R}^p} k\left(\boldsymbol{\beta}, \boldsymbol{\beta}\right) d\boldsymbol{\beta} \ < \ \infty. \tag{2.2}$$

If the trace-class property holds, then the spectrum of K is countable and the corresponding eigenvalues are summable. This in particular implies that K is compact, and the associated Markov chain is geometrically ergodic.

The following theorem shows that the Markov operator K corresponding to the robit DA chain is trace class under very general conditions.

Theorem 1. *The Markov operator K corresponding to the DA Markov chain is trace-class for an arbitrary choice of the design matrix X, sample size n, number of predictors p, degrees of freedom $\nu > 0$, prior mean vector $\boldsymbol{\beta}_a$, and (positive definite) prior precision matrix Σ_a.*

Proof. We shall show that (2.2) holds for the DA Markov chain. The proof is quite lengthy and involved, and we have tried to make it accessible to the reader by highlighting the major steps/milestones.

We begin by fixing our notations. Let $I_A(.)$ be the indicator function of the set A and $\phi\left(x; a, b\right)$ be the univariate normal density evaluated at point x with mean a and variance b. Further, let $\phi_p\left(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ denote the p-variate normal density with mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, evaluated at a vector $\mathbf{x} \in \mathbb{R}^p$. Finally, let $q\left(\omega; a, b\right) = b^a \omega^{a-1} e^{-b\omega}/\Gamma(a)$ be the gamma density evaluated at ω with shape parameter a and rate parameter b.

Note from Section 2.1 of [16] that the joint posterior density of $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{z})$ is given by

$$\begin{aligned}
&\pi\left(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{z} \middle| \mathbf{y}\right) \\
&= \frac{1}{m(\mathbf{y})} \left[\prod_{i=1}^{n} \left\{ I_{\mathbb{R}_+}\left(z_i\right) I_{\{1\}}\left(y_i\right) + I_{\mathbb{R}_-}\left(z_i\right) I_{\{0\}}\left(y_i\right) \right\} \right. \\
&\qquad\qquad \left. \phi\left(z_i; \mathbf{x}_i^T \boldsymbol{\beta}, \frac{1}{\lambda_i}\right) q\left(\lambda_i; \frac{\nu}{2}, \frac{\nu}{2}\right) \right] \times \phi_p\left(\boldsymbol{\beta}; \boldsymbol{\beta}_a, \Sigma_a^{-1}\right) \\
&= \frac{1}{m(\mathbf{y})} \left[\prod_{i=1}^{n} \left\{ I_{\mathbb{R}_+}\left(z_i\right) I_{\{1\}}\left(y_i\right) + I_{\mathbb{R}_-}\left(z_i\right) I_{\{0\}}\left(y_i\right) \right\} \right. \\
&\qquad \left. \times \frac{\sqrt{\lambda_i}}{\sqrt{2\pi}} \exp\left\{ -\frac{\lambda_i}{2} \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}\right)^2 \right\} \times \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \lambda_i^{\frac{\nu}{2}-1} \exp\left\{ -\frac{\nu}{2} \lambda_i \right\} \right] \\
&\qquad\qquad \times (2\pi)^{-\frac{p}{2}} \sqrt{\det\left(\Sigma_a\right)} \ \exp\left[-\frac{1}{2} \left(\boldsymbol{\beta} - \boldsymbol{\beta}_a\right)^T \Sigma_a \left(\boldsymbol{\beta} - \boldsymbol{\beta}_a\right) \right],
\end{aligned}$$

for $\boldsymbol{\lambda} \in \mathbb{R}_+^n, \ \mathbf{z} \in \mathbb{R}^n, \ \boldsymbol{\beta} \in \mathbb{R}^p$.

Step I: A useful linear reparametrization to adjust for the prior mean β_a and derivation of associated conditionals. Consider the following reparametrization

$$\left(\mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\beta}\right) \rightarrow \left(\tilde{\mathbf{z}}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\beta}}\right),$$

using the transformation

$$\tilde{z}_i = z_i - \mathbf{x}_i^T \boldsymbol{\beta}_a \ , \ \text{for all } i = 1, 2, \ldots, n \quad \text{and} \quad \tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{\beta}_a.$$

The absolute value of the Jacobian of this transformation is one, and the joint posterior density of $\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\lambda}, \tilde{\mathbf{z}}\right)$ is given by

$$\begin{aligned}
&\pi\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\lambda}, \tilde{\mathbf{z}} \middle| \mathbf{y}\right) \\
&= \frac{1}{m(\mathbf{y})} \Bigg[\prod_{i=1}^n \left\{ I_{\mathbb{R}_+}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{1\}}\left(y_i\right) + I_{\mathbb{R}_-}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{0\}}\left(y_i\right) \right\} \\
&\qquad \times \frac{\sqrt{\lambda_i}}{\sqrt{2\pi}} \exp\left\{ -\frac{\lambda_i}{2}\left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2 \right\} \times \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \ \lambda_i^{\frac{\nu}{2}-1} \exp\left\{ -\frac{\nu}{2}\lambda_i \right\} \Bigg] \\
&\qquad \times (2\pi)^{-\frac{p}{2}} \sqrt{\det\left(\Sigma_a\right)} \ \exp\left[-\frac{1}{2} \ \tilde{\boldsymbol{\beta}}^T \Sigma_a \tilde{\boldsymbol{\beta}} \right].
\end{aligned} \tag{2.3}$$

Straightforward calculations using (2.3) show that

$$\tilde{\boldsymbol{\beta}} \middle| \boldsymbol{\lambda}, \tilde{\mathbf{z}}, \mathbf{y} \ \sim \ \mathcal{N}_p\left(\left(X^T \Lambda X + \Sigma_a\right)^{-1} X^T \Lambda \tilde{\mathbf{z}} \ , \ \left(X^T \Lambda X + \Sigma_a\right)^{-1} \right),$$

and

$$\begin{aligned}
\pi\left(\tilde{\boldsymbol{\beta}} \middle| \boldsymbol{\lambda}, \tilde{\mathbf{z}}, \mathbf{y}\right) &= (2\pi)^{-\frac{p}{2}} \sqrt{\det\left(X^T \Lambda X + \Sigma_a\right)} \\
&\quad \times \exp\Bigg[-\frac{1}{2} \Bigg\{ \tilde{\boldsymbol{\beta}}^T \left(X^T \Lambda X + \Sigma_a\right) \tilde{\boldsymbol{\beta}} - 2\tilde{\boldsymbol{\beta}}^T X^T \Lambda \tilde{\mathbf{z}} \\
&\qquad + \tilde{\mathbf{z}}^T \Lambda X \left(X^T \Lambda X + \Sigma_a\right)^{-1} X^T \Lambda \tilde{\mathbf{z}} \Bigg\} \Bigg].
\end{aligned} \tag{2.4}$$

It is easy to notice from (2.3) that $(\lambda_i, \tilde{z}_i)$'s are conditionally independent given $\left(\tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$, and moreover,

$$\begin{aligned}
\pi\left(\lambda_i, \tilde{z}_i \middle| \tilde{\boldsymbol{\beta}}, y_i\right) &\propto \left\{ I_{\mathbb{R}_+}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{1\}}\left(y_i\right) + I_{\mathbb{R}_-}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{0\}}\left(y_i\right) \right\} \\
&\quad \times \lambda_i^{\frac{\nu+1}{2}-1} \ \exp\left[-\frac{\lambda_i}{2} \left\{ \nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2 \right\} \right].
\end{aligned} \tag{2.5}$$

Hence, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are conditionally independent given $\left(\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$, and

$$\lambda_i \middle| \tilde{z}_i, \tilde{\boldsymbol{\beta}}, y_i \ \sim \ \text{Gamma}\left(\frac{\nu+1}{2} \ , \ \frac{\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2}{2} \right),$$

for each $i \in \{1, 2, \ldots, n\}$, which implies that

$$\begin{aligned}
\pi\left(\boldsymbol{\lambda}\middle|\tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) &= K_1 \left[\prod_{i=1}^{n} \left(\frac{\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2}{2}\right)^{\frac{\nu+1}{2}}\right] \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu-1}{2}}\right] \\
&\quad \times \exp\left[-\frac{1}{2}\sum_{i=1}^{n} \lambda_i \left(\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2\right)\right] \\
&= K_1' \left[\prod_{i=1}^{n} \left(\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2\right)^{\frac{\nu+1}{2}}\right] \\
&\quad \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu-1}{2}}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n} \lambda_i\right] \\
&\quad \times \exp\left[-\frac{1}{2}\left\{\tilde{\mathbf{z}}^T \Lambda \tilde{\mathbf{z}} - 2\tilde{\boldsymbol{\beta}}^T X^T \Lambda \tilde{\mathbf{z}} + \tilde{\boldsymbol{\beta}}^T X^T \Lambda X \tilde{\boldsymbol{\beta}}\right\}\right],
\end{aligned} \tag{2.6}$$

where K_1 and K_1' are appropriate constants.

To find $\pi\left(\tilde{\mathbf{z}}\middle|\tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$, we use (2.5) to get for each $i \in \{1, 2, \ldots, n\}$,

$$\begin{aligned}
&\pi\left(\tilde{z}_i\middle|\tilde{\boldsymbol{\beta}}, y_i\right) \\
&= \int_{\mathbb{R}_+} \pi\left(\lambda_i, \tilde{z}_i\middle|\tilde{\boldsymbol{\beta}}, y_i\right) d\lambda_i \\
&= C_{\tilde{\boldsymbol{\beta}}, y_i} \times \left[I_{\mathbb{R}_+}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{1\}}(y_i) + I_{\mathbb{R}_-}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{0\}}(y_i)\right] \\
&\qquad \times \int_{\mathbb{R}_+} \lambda_i^{\frac{\nu+1}{2}-1} \exp\left[-\frac{\lambda_i}{2}\left\{\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2\right\}\right] d\lambda_i \\
&\left[\text{where, } C_{\tilde{\boldsymbol{\beta}}, y_i} \text{ is a constant which depends on } \tilde{\boldsymbol{\beta}}, y_i\right] \\
&= C_{\tilde{\boldsymbol{\beta}}, y_i} \times \left[I_{\mathbb{R}_+}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{1\}}(y_i) + I_{\mathbb{R}_-}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{0\}}(y_i)\right] \\
&\qquad \times \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\left(\frac{\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2}{2}\right)^{\frac{\nu+1}{2}}} \\
&= C'_{\tilde{\boldsymbol{\beta}}, y_i} \times \left[I_{\mathbb{R}_+}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{1\}}(y_i) + I_{\mathbb{R}_-}\left(\tilde{z}_i + \mathbf{x}_i^T \boldsymbol{\beta}_a\right) I_{\{0\}}(y_i)\right] \\
&\qquad \times \left(1 + \frac{\left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}},
\end{aligned} \tag{2.7}$$

where $C'_{\tilde{\boldsymbol{\beta}}, y_i}$ is the product of all constant terms that are free of $\tilde{z}_i$. We conclude from (2.3) and (2.7) that conditional on $\left(\tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$, $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n$ are independent with $\tilde{z}_i\middle|\tilde{\boldsymbol{\beta}}, \mathbf{y}$ following a truncated t distribution with location $\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}$, scale 1 and degrees of freedom ν that is truncated left at $-\mathbf{x}_i^T \boldsymbol{\beta}_a$ if $y_i = 1$ and truncated right at $-\mathbf{x}_i^T \boldsymbol{\beta}_a$ if $y_i = 0$.

Now, if we denote $t_\nu(\mu, 1)$ to be the univariate Student's t distribution with location μ, scale 1 and degrees of freedom ν, and F_ν to be the cdf of the $t_\nu(0,1)$ distribution, then for $y_i = 0$,

$$\pi\left(\tilde{z}_i \middle| \tilde{\boldsymbol{\beta}}, y_i\right) = \frac{K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(\mathbf{x}_i^T\tilde{\boldsymbol{\beta}},\ 1\right) \le -\mathbf{x}_i^T\boldsymbol{\beta}_a\right)} = \frac{K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,\ 1\right) \le -\mathbf{x}_i^T\left(\tilde{\boldsymbol{\beta}}+\boldsymbol{\beta}_a\right)\right)} = \frac{K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{1-F_\nu\left(\mathbf{x}_i^T\left(\tilde{\boldsymbol{\beta}}+\boldsymbol{\beta}_a\right)\right)}, \tag{2.8}$$

and for $y_i = 1$,

$$\pi\left(\tilde{z}_i \middle| \tilde{\boldsymbol{\beta}}, y_i\right) = \frac{K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(\mathbf{x}_i^T\tilde{\boldsymbol{\beta}},\ 1\right) \ge -\mathbf{x}_i^T\boldsymbol{\beta}_a\right)} = \frac{K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,\ 1\right) \ge -\mathbf{x}_i^T\left(\tilde{\boldsymbol{\beta}}+\boldsymbol{\beta}_a\right)\right)} = \frac{K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{F_\nu\left(\mathbf{x}_i^T\left(\tilde{\boldsymbol{\beta}}+\boldsymbol{\beta}_a\right)\right)}. \tag{2.9}$$

Combining (2.8) and (2.9), we have for any y_i,

$$\pi\left(\tilde{z}_i \middle| \tilde{\boldsymbol{\beta}}, y_i\right) = K_2\left(1+\frac{\left(\tilde{z}_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \times \left\{\frac{1}{F_\nu\left(\mathbf{x}_i^T\left(\tilde{\boldsymbol{\beta}}+\boldsymbol{\beta}_a\right)\right)}\right\}^{y_i}\left\{\frac{1}{1-F_\nu\left(\mathbf{x}_i^T\left(\tilde{\boldsymbol{\beta}}+\boldsymbol{\beta}_a\right)\right)}\right\}^{1-y_i},$$

which implies,

$$\pi\left(\tilde{\mathbf{z}} \middle| \tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$$

$$= K_2^n \, \nu^{\frac{n(\nu+1)}{2}} \prod_{i=1}^{n} \left[\left(\nu + \left(\tilde{z}_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}\right)^2\right)^{-\frac{\nu+1}{2}} \left\{ \frac{1}{F_\nu\left(\mathbf{x}_i^T \left(\tilde{\boldsymbol{\beta}} + \boldsymbol{\beta}_a\right)\right)} \right\}^{y_i} \right.$$
$$\left. \times \left\{ \frac{1}{1 - F_\nu\left(\mathbf{x}_i^T \left(\tilde{\boldsymbol{\beta}} + \boldsymbol{\beta}_a\right)\right)} \right\}^{1-y_i} \right]. \tag{2.10}$$

Let $S := \{i : y_i = 0\}$. Then, $S^c = \{i : \; y_i = 1\}$. Also, let

$$\tilde{k}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}') = \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^n} \pi\left(\tilde{\boldsymbol{\beta}}' \middle| \boldsymbol{\lambda}, \tilde{\mathbf{z}}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \middle| \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) \pi\left(\tilde{\mathbf{z}} \middle| \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) d\tilde{\mathbf{z}} \; d\boldsymbol{\lambda},$$

denote the transition density of the marginal $\tilde{\boldsymbol{\beta}}$-chain which tracks the linearly transformed version $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} - \boldsymbol{\beta}_a$ of $\boldsymbol{\beta}$. Using (2.1), (2.4), (2.6) and (2.10), we get the following form for the integral I in (2.2) under the new parametrization.

$$\begin{aligned}
I &= \int_{\mathbb{R}^p} \tilde{k}\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}\right) d\tilde{\boldsymbol{\beta}} \\
&= \int_{\mathbb{R}^p} \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^n} \pi\left(\tilde{\boldsymbol{\beta}} \middle| \boldsymbol{\lambda}, \tilde{\mathbf{z}}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \middle| \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) \pi\left(\tilde{\mathbf{z}} \middle| \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) d\tilde{\mathbf{z}} \; d\boldsymbol{\lambda} \; d\tilde{\boldsymbol{\beta}} \\
&= \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} \pi\left(\tilde{\boldsymbol{\beta}} \middle| \boldsymbol{\lambda}, \tilde{\mathbf{z}}, \mathbf{y}\right) \pi\left(\boldsymbol{\lambda} \middle| \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) \pi\left(\tilde{\mathbf{z}} \middle| \tilde{\boldsymbol{\beta}}, \mathbf{y}\right) d\tilde{\mathbf{z}} \; d\tilde{\boldsymbol{\beta}} \; d\boldsymbol{\lambda} \\
&= C_0 \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} \sqrt{\det\left(X^T \Lambda X + \Sigma_a\right)} \; \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu-1}{2}}\right] \times \; \exp\left[-\frac{\nu}{2} \sum_{i=1}^n \lambda_i\right] \\
&\qquad \times \prod_{i \in S} \left[\frac{1}{1 - F_\nu\left(\mathbf{x}_i^T \left(\tilde{\boldsymbol{\beta}} + \boldsymbol{\beta}_a\right)\right)}\right] \times \prod_{i \in S^c} \left[\frac{1}{F_\nu\left(\mathbf{x}_i^T \left(\tilde{\boldsymbol{\beta}} + \boldsymbol{\beta}_a\right)\right)}\right] \\
&\qquad \times \; \exp\left[-\frac{1}{2} \left\{ \tilde{\boldsymbol{\beta}}^T \left(X^T \Lambda X + \Sigma_a\right) \tilde{\boldsymbol{\beta}} \; - \; 2 \tilde{\boldsymbol{\beta}}^T X^T \Lambda \tilde{\mathbf{z}} \right.\right. \\
&\qquad\qquad \left.\left. + \; \tilde{\mathbf{z}}^T \Lambda X \left(X^T \Lambda X + \Sigma_a\right)^{-1} X^T \Lambda \tilde{\mathbf{z}} \right\}\right] \\
&\qquad \times \; \exp\left[-\frac{1}{2} \left\{ \tilde{\mathbf{z}}^T \Lambda \tilde{\mathbf{z}} - 2 \tilde{\boldsymbol{\beta}}^T X^T \Lambda \tilde{\mathbf{z}} \right.\right. \\
&\qquad\qquad \left.\left. + \; \tilde{\boldsymbol{\beta}}^T X^T \Lambda X \tilde{\boldsymbol{\beta}} \right\}\right] \; d\tilde{\mathbf{z}} \; d\tilde{\boldsymbol{\beta}} \; d\boldsymbol{\lambda}.
\end{aligned} \tag{2.11}$$

Here, C_0 denotes the product of all constant terms (independent of $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\lambda}$, and $\tilde{\mathbf{z}}$) appearing in the conditional densities $\pi\left(\tilde{\boldsymbol{\beta}} \middle| \boldsymbol{\lambda}, \tilde{\mathbf{z}}, \mathbf{y}\right)$, $\pi\left(\boldsymbol{\lambda} \middle| \tilde{\mathbf{z}}, \tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$, and $\pi\left(\tilde{\mathbf{z}} \middle| \tilde{\boldsymbol{\beta}}, \mathbf{y}\right)$.

Step II: Another reparametrization to adjust for the prior precision matrix Σ_a. Now, let us define $\boldsymbol{\theta} = \Sigma_a^{1/2} \tilde{\boldsymbol{\beta}}$, $W = X \Sigma_a^{-1/2}$, and $\tilde{\mathbf{c}} = \Sigma_a^{1/2} \boldsymbol{\beta}_a$. Absolute value of the Jacobian of the transformation $\tilde{\boldsymbol{\beta}} \to \boldsymbol{\theta}$ is $\{\det\left(\Sigma_a\right)\}^{-1/2} >$

0. Therefore, the right hand side of (2.11), after this further reparametrization becomes

$$
\begin{aligned}
I = C_0 \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} & \sqrt{\det\left(W^T \Lambda W + I_p\right)} \; \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu-1}{2}}\right] \times \; \exp\left[-\frac{\nu}{2}\sum_{i=1}^n \lambda_i\right] \\
& \times \prod_{i \in S}\left[\frac{1}{1 - F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \times \prod_{i \in S^c}\left[\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \\
& \times \; \exp\Bigg[-\frac{1}{2}\Bigg\{\boldsymbol{\theta}^T\left(W^T \Lambda W + I_p\right)\boldsymbol{\theta} \; - \; 2\boldsymbol{\theta}^T W^T \Lambda \tilde{\mathbf{z}} \\
& \qquad\qquad + \; \tilde{\mathbf{z}}^T \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda \tilde{\mathbf{z}}\Bigg\}\Bigg] \\
& \times \; \exp\Bigg[-\frac{1}{2}\left\{\tilde{\mathbf{z}}^T \Lambda \tilde{\mathbf{z}} - 2\boldsymbol{\theta}^T W^T \Lambda \tilde{\mathbf{z}}\right. \\
& \qquad\qquad \left. + \; \boldsymbol{\theta}^T W^T \Lambda W \boldsymbol{\theta}\right\}\Bigg] \; d\tilde{\mathbf{z}} \; d\boldsymbol{\theta} \; d\boldsymbol{\lambda} \\
= C_0 \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} & \sqrt{\det\left(W^T \Lambda W + I_p\right)} \; \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu-1}{2}}\right] \times \; \exp\left[-\frac{\nu}{2}\sum_{i=1}^n \lambda_i\right] \\
& \times \prod_{i \in S}\left[\frac{1}{1 - F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \times \prod_{i \in S^c}\left[\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \\
& \times \; \exp\left[-\frac{1}{2}\left\{\boldsymbol{\theta}^T\left(2W^T \Lambda W + I_p\right)\boldsymbol{\theta}\right\}\right] \\
& \times \Bigg(\int_{\mathbb{R}^n} \exp\Big[2\boldsymbol{\theta}^T W^T \Lambda \tilde{\mathbf{z}} \\
& \qquad - \; \frac{1}{2}\,\tilde{\mathbf{z}}^T\left(\Lambda + \Lambda W\left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)\tilde{\mathbf{z}}\Big] \, d\tilde{\mathbf{z}}\Bigg) \, d\boldsymbol{\theta} \; d\boldsymbol{\lambda}.
\end{aligned}
\tag{2.12}
$$

Step III: An upper bound for the innermost $\tilde{\mathbf{z}}$ integral in (2.12). We now derive an upper bound for the innermost integral in (2.12). Note that

$$
\begin{aligned}
& \int_{\mathbb{R}^n} \exp\left[2\boldsymbol{\theta}^T W^T \Lambda \tilde{\mathbf{z}} \; - \; \frac{1}{2}\,\tilde{\mathbf{z}}^T\left(\Lambda + \Lambda W\left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)\tilde{\mathbf{z}}\right] \; d\tilde{\mathbf{z}} \\
& = \; \exp\left[\frac{1}{2}\; 4\boldsymbol{\theta}^T W^T \Lambda\left(\Lambda + \Lambda W\left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \Lambda W \boldsymbol{\theta}\right] \times (C_1)^{-1} \\
& \times \int_{\mathbb{R}^n} C_1 \exp\left[-\frac{1}{2}\left(\tilde{\mathbf{z}} - \mathbf{a}_1^\star\right)^T\left(\Lambda + \Lambda W\left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)\left(\tilde{\mathbf{z}} - \mathbf{a}_1^\star\right)\right] \; d\tilde{\mathbf{z}} \\
& \qquad \left[\text{where, } \mathbf{a}_1^\star = 2\left(\Lambda + \Lambda W\left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \Lambda W \boldsymbol{\theta}\right]
\end{aligned}
$$

$$= C_1^{-1} \exp\left[\frac{1}{2}\, 4\boldsymbol{\theta}^T W^T \Lambda \left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \Lambda W \boldsymbol{\theta}\right], \tag{2.13}$$

where $C_1 = (2\pi)^{-n/2}\left\{\det\left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)\right\}^{1/2}$. The last equality follows from the fact that the integrand is a normal density. However,

$$\begin{aligned} C_1 &= (2\pi)^{-n/2}\left\{\det\left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)\right\}^{1/2} \\ &\geq (2\pi)^{-n/2}\left\{\det\left(\Lambda\right)\right\}^{1/2} \\ &= (2\pi)^{-n/2}\sqrt{\prod_{i=1}^{n} \lambda_i}\;. \end{aligned} \tag{2.14}$$

From (2.13) and (2.14), we have an upper bound for the inner integral in (2.12) as follows

$$\begin{aligned} &\int_{\mathbb{R}^n} \exp\left[2\boldsymbol{\theta}^T W^T \Lambda \tilde{\mathbf{z}} \;-\; \frac{1}{2}\,\tilde{\mathbf{z}}^T \left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right) \tilde{\mathbf{z}}\right]\; d\tilde{\mathbf{z}} \\ &\leq (2\pi)^{n/2} \prod_{i=1}^{n} \lambda_i^{-1/2} \\ &\quad \times \exp\left[\frac{1}{2}\, 4\boldsymbol{\theta}^T W^T \Lambda \left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \Lambda W \boldsymbol{\theta}\right]. \end{aligned} \tag{2.15}$$

Combining (2.12) and (2.15), we get

$$\begin{aligned} I \leq C_0\, (2\pi)^{n/2} &\int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \sqrt{\det\left(W^T \Lambda W + I_p\right)} \;\times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n} \lambda_i\right] \\ &\times \prod_{i \in S}\left[\frac{1}{1 - F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \times \prod_{i \in S^c}\left[\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \\ &\times \exp\Bigg[-\frac{1}{2}\Bigg\{\boldsymbol{\theta}^T \left(2W^T \Lambda W + I_p\right) \boldsymbol{\theta} \\ &\qquad - 4\boldsymbol{\theta}^T W^T \Lambda \left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \\ &\qquad \times \Lambda W \boldsymbol{\theta}\Bigg\}\Bigg]\; d\boldsymbol{\theta}\; d\boldsymbol{\lambda} \\ = C_2 &\int_{\mathbb{R}^p} \int_{\mathbb{R}_+^n} \sqrt{\det\left(W^T \Lambda W + I_p\right)} \;\times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n} \lambda_i\right] \\ &\times \prod_{i \in S}\left[\frac{1}{1 - F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \times \prod_{i \in S^c}\left[\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)}\right] \end{aligned}$$

$$\times \exp\left[-\frac{1}{2}\Big\{G(\boldsymbol{\theta}, \boldsymbol{\lambda})\Big\}\right] \; d\boldsymbol{\theta} \; d\boldsymbol{\lambda}, \tag{2.16}$$

where $C_2 = C_0\,(2\pi)^{n/2}$ and

$$\begin{aligned} G(\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \boldsymbol{\theta}^T\left(2W^T\Lambda W + I_p\right)\boldsymbol{\theta} \\ &\qquad - 4\boldsymbol{\theta}^T W^T\Lambda\left(\Lambda + \Lambda W\left(W^T\Lambda W + I_p\right)^{-1} W^T\Lambda\right)^{-1}\Lambda W\boldsymbol{\theta} \\ &= \boldsymbol{\theta}^T\Big[\left(2W^T\Lambda W + I_p\right) \\ &\qquad -4W^T\Lambda\left(\Lambda + \Lambda W\left(W^T\Lambda W + I_p\right)^{-1} W^T\Lambda\right)^{-1}\Lambda W\Big]\boldsymbol{\theta}. \end{aligned} \tag{2.17}$$

Step IV: An upper bound for the products involving the cdf F_ν. We now target the product terms in the integrand involving the t-cdf F_ν. Note that for $i \in S$, if $\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}}) \le 0$, then

$$F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right) \le F_\nu(0) = \frac{1}{2} \implies \frac{1}{1 - F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right)} \le 2, \tag{2.18}$$

and if $\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}}) > 0$, then by Lemma A.1 in Appendix A we have

$$\frac{1}{1 - F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right)} \le \frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta} + \mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2 + \nu\right)^{\frac{\nu}{2}}}{\kappa}. \tag{2.19}$$

From (2.18) and (2.19), we have for any $i \in S$,

$$\begin{aligned} \frac{1}{1 - F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right)} &\le \max\left\{2\,,\; \frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta} + \mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2 + \nu\right)^{\frac{\nu}{2}}}{\kappa}\right\} \\ &\le \left(2 + \frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta} + \mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2 + \nu\right)^{\frac{\nu}{2}}}{\kappa}\right). \end{aligned} \tag{2.20}$$

Similarly for $i \in S^c$, if $\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}}) \ge 0$, then

$$\begin{aligned} &F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right) \ge F_\nu(0) = \frac{1}{2} \\ \implies &\frac{1}{F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right)} \le 2, \end{aligned} \tag{2.21}$$

and if $\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}}) < 0$, i.e., $-\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}}) > 0$, then by Lemma A.1 in Appendix A we have

$$\frac{1}{F_\nu\left(\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right)} \;=\; \frac{1}{1 - F_\nu\left(-\mathbf{w}_i^T(\boldsymbol{\theta} + \tilde{\mathbf{c}})\right)} \;\le\; \frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta} + \mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2 + \nu\right)^{\frac{\nu}{2}}}{\kappa}. \tag{2.22}$$

From (2.21) and (2.22), we have for any $i \in S^c$,

$$\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\right)} \leq \max\left\{2\ ,\ \frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta}+\mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2+\nu\right)^{\frac{\nu}{2}}}{\kappa}\right\}$$
$$\leq \left(2+\frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta}+\mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2+\nu\right)^{\frac{\nu}{2}}}{\kappa}\right). \tag{2.23}$$

Finally from (2.20) and (2.23), we have

$$\prod_{i\in S}\left[\frac{1}{1-F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\right)}\right]\times\prod_{i\in S^c}\left[\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\right)}\right]$$
$$\leq \prod_{i=1}^{n}\left(2+\frac{\left(\left(\mathbf{w}_i^T\boldsymbol{\theta}+\mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2+\nu\right)^{\frac{\nu}{2}}}{\kappa}\right). \tag{2.24}$$

Note that

$$\begin{aligned}\left(\mathbf{w}_i^T\boldsymbol{\theta}+\mathbf{w}_i^T\tilde{\mathbf{c}}\right)^2 &= \left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T\mathbf{w}_i\mathbf{w}_i^T\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\\ &\leq \left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T\left(\sum_{i=1}^{n}\mathbf{w}_i\mathbf{w}_i^T\right)\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\\ &= \left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T W^TW\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right),\end{aligned} \tag{2.25}$$

for every $1 \leq i \leq n$. It follows from (2.24), (2.25), and the c_r-inequality that

$$\begin{aligned}&\prod_{i\in S}\left[\frac{1}{1-F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\right)}\right]\times\prod_{i\in S^c}\left[\frac{1}{F_\nu\left(\mathbf{w}_i^T\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\right)}\right]\\
&\leq \prod_{i=1}^{n}\left(2+\frac{\left(\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T W^TW\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)+\nu\right)^{\frac{\nu}{2}}}{\kappa}\right)\\
&= \left(2+\frac{\left(\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T W^TW\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)+\nu\right)^{\frac{\nu}{2}}}{\kappa}\right)^n\\
&\leq\ 2^n\left[2^n\ +\ \frac{\left(\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T W^TW\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)+\nu\right)^{\frac{n\nu}{2}}}{\kappa^n}\right]\\
&\leq\ 2^n\left[2^n\ +\ \frac{2^{n\nu/2}\left\{\left(\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)^T W^TW\left(\boldsymbol{\theta}+\tilde{\mathbf{c}}\right)\right)^{\frac{n\nu}{2}}+\nu^{n\nu/2}\right\}}{\kappa^n}\right]\end{aligned}$$

$$= C_3 + C_4 \left((\boldsymbol{\theta} + \tilde{\mathbf{c}})^T W^T W (\boldsymbol{\theta} + \tilde{\mathbf{c}}) \right)^{\frac{n\nu}{2}}$$
$$\leq C_3 + C_6 \left((\boldsymbol{\theta} + \tilde{\mathbf{c}})^T (\boldsymbol{\theta} + \tilde{\mathbf{c}}) \right)^{\frac{n\nu}{2}}, \tag{2.26}$$

where $C_3 = 2^{2n} + \frac{2^n (2\nu)^{n\nu/2}}{\kappa^n}$, $C_4 = \frac{2^n 2^{n\nu/2}}{\kappa^n}$, C_5 denotes the largest eigenvalue of $W^T W$, and $C_6 = C_4 C_5^{\frac{n\nu}{2}}$.

Let $\lambda_{\max}$ denote the largest diagonal entry of the diagonal matrix Λ. Using (2.16), (2.26), $\left(W^T \Lambda W + I_p\right) \preceq \max(1, \lambda_{\max}) \left(W^T W + I_p\right)$, and Fubini's theorem, it follows that

$$\begin{aligned} I \leq\ & C_2 \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \max\left\{1, \lambda_{\max}^{p/2}\right\} \sqrt{\det\left(W^T W + I_p\right)} \\ & \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2} \sum_{i=1}^n \lambda_i\right] \\ & \times \left[C_3 + C_6 \left((\boldsymbol{\theta} + \tilde{\mathbf{c}})^T (\boldsymbol{\theta} + \tilde{\mathbf{c}}) \right)^{\frac{n\nu}{2}}\right] \\ & \times \exp\left[-\frac{1}{2} \left\{ G(\boldsymbol{\theta}, \boldsymbol{\lambda}) \right\}\right] d\boldsymbol{\theta}\, d\boldsymbol{\lambda} \\ = \ & C_7 \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2} \sum_{i=1}^n \lambda_i\right] \\ & \times \left[C_3 + C_6 \left((\boldsymbol{\theta} + \tilde{\mathbf{c}})^T (\boldsymbol{\theta} + \tilde{\mathbf{c}}) \right)^{\frac{n\nu}{2}}\right] \\ & \times \exp\left[-\frac{1}{2} \left\{ G(\boldsymbol{\theta}, \boldsymbol{\lambda}) \right\}\right] d\boldsymbol{\theta}\, d\boldsymbol{\lambda}, \end{aligned} \tag{2.27}$$

where $C_7 = C_2 \sqrt{\det\left(W^T W + I_p\right)}$ is a constant term free of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$.

Step V: Showing $G(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is positive definite quadratic form in $\boldsymbol{\theta}$. In order to show the finiteness of the upper bound for I in (2.27), we will first prove that $G(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is a positive definite quadratic form in $\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \mathbb{R}^p$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^n$. For that, it is enough to show by (2.17) that the matrix

$$\left[\left(2W^T \Lambda W + I_p\right) - 4W^T \Lambda \left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \Lambda W\right],$$

is a positive definite matrix for all $\boldsymbol{\lambda} \in \mathbb{R}_+^n$. We show this by working out the spectral decomposition of this matrix separately in the low and high-dimensional settings.

Low-dimensional setting: When $n \geq p$

$$\left(2W^T \Lambda W + I_p\right) - 4W^T \Lambda \left(\Lambda + \Lambda W \left(W^T \Lambda W + I_p\right)^{-1} W^T \Lambda\right)^{-1} \Lambda W$$

$$
\begin{aligned}
&= \left(2W^T\Lambda W + I_p\right) \\
&\qquad - 4W^T\Lambda^{1/2}\left(I_n + \Lambda^{1/2}W\left(W^T\Lambda W + I_p\right)^{-1}W^T\Lambda^{1/2}\right)^{-1}\Lambda^{1/2}W \\
&= \left(2A^TA + I_p\right) - 4A^T\left(I_n + A\left(A^TA + I_p\right)^{-1}A^T\right)^{-1}A, \qquad (2.28)
\end{aligned}
$$

where $A = \Lambda^{1/2}W$. Now, by the Singular Value Decomposition, A can be written as

$$A_{n\times p} = U_{n\times p}D_{p\times p}V^T_{p\times p}, \qquad (2.29)$$

where U is a semi-orthogonal matrix i.e. $U^TU = I_p$, V is an orthogonal matrix i.e. $V^TV = VV^T = I_p$, and D is a diagonal matrix with singular values of A being the diagonal entries. Since $U_{n\times p}$ is a semi-orthogonal matrix with full column rank p, there exists a matrix $U_{0\ n\times(n-p)}$ such that the matrix $\begin{bmatrix}U_{n\times p} & U_{0\ n\times(n-p)}\end{bmatrix}$ becomes an orthogonal matrix, i.e.,

$$\begin{bmatrix}U & U_0\end{bmatrix}\begin{bmatrix}U^T \\ U_0^T\end{bmatrix} = \begin{bmatrix}U^T \\ U_0^T\end{bmatrix}\begin{bmatrix}U & U_0\end{bmatrix} = I_n,$$

which implies that

$$U_0^TU = 0_{(n-p)\times p}\ ,\ U^TU_0 = 0_{p\times(n-p)}.$$

Using $A = UDV^T$ in (2.28), and standard matrix algebra leveraging the various orthogonality properties discussed above, we get

$$
\begin{aligned}
&\left(2W^T\Lambda W + I_p\right) - 4W^T\Lambda\left(\Lambda + \Lambda W\left(W^T\Lambda W + I_p\right)^{-1}W^T\Lambda\right)^{-1}\Lambda W \\
&= \left(2A^TA + I_p\right) - 4A^T\left(I_n + A\left(A^TA + I_p\right)^{-1}A^T\right)^{-1}A \\
&= \left(2VD^2V^T + I_p\right) \\
&\quad - 4VDU^T\left(\begin{bmatrix}U & U_0\end{bmatrix}\begin{bmatrix}U^T \\ U_0^T\end{bmatrix}\right. \\
&\qquad\qquad + \begin{bmatrix}U & U_0\end{bmatrix}\begin{bmatrix}U^T \\ U_0^T\end{bmatrix}UDV^T\left(VD^2V^T + I_p\right)^{-1} \\
&\qquad\qquad\qquad \left.\times VDU^T\begin{bmatrix}U & U_0\end{bmatrix}\begin{bmatrix}U^T \\ U_0^T\end{bmatrix}\right)^{-1}UDV^T \\
&= \left(2VD^2V^T + I_p\right) \\
&\quad - 4VDU^T\left(\begin{bmatrix}U & U_0\end{bmatrix}\begin{bmatrix}U^T \\ U_0^T\end{bmatrix}\right. \\
&\qquad\qquad \left. + \begin{bmatrix}U & U_0\end{bmatrix}\begin{bmatrix}DV^T\left(VD^2V^T + I_p\right)^{-1}VD & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}U^T \\ U_0^T\end{bmatrix}\right)^{-1}UDV^T \\
&= \left(2VD^2V^T + I_p\right)
\end{aligned}
$$

$$
\begin{aligned}
&- 4VDU^T \begin{bmatrix} U & U_0 \end{bmatrix} \left(I_n + \begin{bmatrix} DV^T \left(VD^2V^T + I_p\right)^{-1} VD & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \\
&\qquad \begin{bmatrix} U^T \\ U_0^T \end{bmatrix} UDV^T \\
&= \left(2VD^2V^T + I_p\right) \\
&\quad - 4VD \begin{bmatrix} I_p & 0 \end{bmatrix} \begin{bmatrix} \left(I_p + DV^T \left(VD^2V^T + I_p\right)^{-1} VD \right)^{-1} & 0 \\ 0 & I_{n-p} \end{bmatrix} \\
&\qquad \begin{bmatrix} I_p \\ 0 \end{bmatrix} DV^T \\
&= \left(2VD^2V^T + I_p\right) - 4VD \left(I_p + DV^T \left(VD^2V^T + I_p\right)^{-1} VD \right)^{-1} DV^T \\
&= \left(2VD^2V^T + VV^T\right) - 4VD \left(I_p + D \left(D^2 + I_p\right)^{-1} D \right)^{-1} DV^T \\
&= V \left(2D^2 + I_p\right) V^T - 4VD \left(I_p + \frac{D^2}{D^2 + I_p} \right)^{-1} DV^T \\
&= V \left[\frac{I_p}{2D^2 + I_p} \right] V^T. \qquad (2.30)
\end{aligned}
$$

<u>High-dimensional setting:</u> When $n < p$, the Singular Value Decomposition of $A = \Lambda^{1/2} W$ can be written as

$$A_{n\times p} = V_{n\times n} D_{n\times n} U^T_{n\times p},$$

where $U_{p\times n}$ is a semi-orthogonal matrix i.e. $U^T U = I_n$, V is an orthogonal matrix i.e. $V^T V = VV^T = I_n$, and D is a diagonal matrix with singular values of A being the diagonal entries. Since $U_{p\times n}$ is a semi-orthogonal matrix with full column rank n, there exists a matrix $U_{0\ p\times(p-n)}$ such that the matrix $\begin{bmatrix} U_{p\times n} & U_{0\ p\times(p-n)} \end{bmatrix}$ becomes an orthogonal matrix, i.e.,

$$\begin{bmatrix} U & U_0 \end{bmatrix} \begin{bmatrix} U^T \\ U_0^T \end{bmatrix} = \begin{bmatrix} U^T \\ U_0^T \end{bmatrix} \begin{bmatrix} U & U_0 \end{bmatrix} = I_p,$$

which implies that

$$U_0^T U = 0_{(p-n)\times n}\ ,\ U^T U_0 = 0_{n\times(p-n)}.$$

Again, using $A = VDU^T$ in (2.28), and standard matrix algebra leveraging the various orthogonality properties discussed above, we get

$$
\begin{aligned}
&\left(2W^T\Lambda W + I_p\right) - 4W^T\Lambda \left(\Lambda + \Lambda W \left(W^T\Lambda W + I_p\right)^{-1} W^T\Lambda \right)^{-1} \Lambda W \\
&= \left(2A^TA + I_p\right) - 4A^T \left(I_n + A \left(A^TA + I_p\right)^{-1} A^T \right)^{-1} A
\end{aligned}
$$

$$
\begin{aligned}
&= \left(2UD^2U^T + I_p\right) - 4UD\left(I_n + DU^T\left(UD^2U^T + I_p\right)^{-1}UD\right)^{-1}DU^T \\
&= \left(2UD^2U^T + I_p\right) \\
&\qquad - 4UD\left(I_n + DU^T\left(UD^2U^T + \begin{bmatrix} U & U_0 \end{bmatrix}\begin{bmatrix} U^T \\ U_0^T \end{bmatrix}\right)^{-1}UD\right)^{-1}DU^T \\
&= \left(2UD^2U^T + I_p\right) \\
&\qquad - 4UD\left(I_n + DU^T\left(U\left(D^2 + I_n\right)U^T + U_0U_0^T\right)^{-1}UD\right)^{-1}DU^T \\
&= \left(2UD^2U^T + I_p\right) \\
&\qquad - 4UD\left(I_n + DU^T\left(U\left(D^2 + I_n\right)^{-1}U^T + U_0U_0^T\right)UD\right)^{-1}DU^T \\
&= U\left(2D^2 + I_n\right)U^T + U_0U_0^T - U\left[4D^2\left(D^2 + I_n\right)\left(2D^2 + I_n\right)^{-1}\right]U^T \\
&= U\left[\left(2D^2 + I_n\right) - \frac{4D^2\left(D^2 + I_n\right)}{2D^2 + I_n}\right]U^T + U_0U_0^T \\
&= U\left[\frac{I_n}{2D^2 + I_n}\right]U^T + U_0U_0^T. \qquad (2.31)
\end{aligned}
$$

Now, if we denote

$$
\Omega(\Lambda) := \left(2W^T\Lambda W + I_p\right) - 4W^T\Lambda\left(\Lambda + \Lambda W\left(W^T\Lambda W + I_p\right)^{-1}W^T\Lambda\right)^{-1}\Lambda W,
$$

then from (2.30) and (2.31), it follows that

$$
\Omega(\Lambda) = \begin{cases} V\left[\frac{I_p}{2D^2 + I_p}\right]V^T & \text{if } n \geq p \\ U\left[\frac{I_n}{2D^2 + I_n}\right]U^T + U_0U_0^T & \text{if } n < p. \end{cases} \qquad (2.32)
$$

Then, clearly $\Omega(\Lambda)$ is a positive definite matrix for both the cases $n \geq p$ and $n < p$, and for all $\boldsymbol{\lambda} \in \mathbb{R}_+^n$. This implies that $G(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \boldsymbol{\theta}^T\Omega(\Lambda)\boldsymbol{\theta}$ is a positive definite quadratic form in $\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \mathbb{R}^p$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^n$. Moreover,

$$
\Sigma(\Lambda) := \Omega(\Lambda)^{-1} = \begin{cases} V\left(2D^2 + I_p\right)V^T & \text{if } n \geq p \\ U\left(2D^2 + I_n\right)U^T + U_0U_0^T & \text{if } n < p. \end{cases} \qquad (2.33)
$$

Now, from (2.27) we get

$$
\begin{aligned}
I \;\leq\; C_7 \int_{\mathbb{R}_+^n}\int_{\mathbb{R}^p} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^n \lambda_i\right] \\
\times \left[C_3 \;+\; C_6\left(\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)^T\left(\boldsymbol{\theta} + \tilde{\mathbf{c}}\right)\right)^{\frac{n\nu}{2}}\right]
\end{aligned}
$$

$$\times \exp\left[-\frac{1}{2}\Big\{G(\boldsymbol{\theta},\boldsymbol{\lambda})\Big\}\right] \, d\boldsymbol{\theta} \, d\boldsymbol{\lambda}$$

$$= C_7 \int_{\mathbb{R}_+^n} \int_{\mathbb{R}^p} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^n \lambda_i\right]$$
$$\times \left[C_3 \ + \ C_6 \left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]$$
$$\times \exp\left[-\frac{1}{2} \, \boldsymbol{\theta}^T \Omega(\Lambda) \boldsymbol{\theta}\right] \, d\boldsymbol{\theta} \, d\boldsymbol{\lambda}$$
$$= C_7 \int_{\mathbb{R}_+^n} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^n \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^n \lambda_i\right]$$
$$\times \left(\int_{\mathbb{R}^p} \left[C_3 \ + \ C_6 \left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]\right.$$
$$\left.\times \exp\left[-\frac{1}{2} \, \boldsymbol{\theta}^T \Omega(\Lambda) \boldsymbol{\theta}\right] \, d\boldsymbol{\theta}\right) \, d\boldsymbol{\lambda}. \tag{2.34}$$

Step VI: An upper bound for the inner integral in (2.34). We now derive an upper bound for the inner integral in (2.34) using properties of the multivariate normal distribution. Note that

$$\int_{\mathbb{R}^p} \left[C_3 \ + \ C_6 \left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right] \exp\left[-\frac{1}{2} \, \boldsymbol{\theta}^T \Omega(\Lambda) \boldsymbol{\theta}\right] \, d\boldsymbol{\theta}$$
$$= (2\pi)^{p/2} \sqrt{\det(\Sigma(\Lambda))}$$
$$\times \int_{\mathbb{R}^p} (2\pi)^{-p/2} \, \det(\Sigma(\Lambda))^{-1/2}$$
$$\times \left[C_3 \ + \ C_6 \left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]$$
$$\times \ \exp\left[-\frac{1}{2} \, \boldsymbol{\theta}^T \Omega(\Lambda) \boldsymbol{\theta}\right] \, d\boldsymbol{\theta}$$
$$= (2\pi)^{p/2} \sqrt{\det(\Sigma(\Lambda))}$$
$$\times \ E_{\mathcal{N}_p(\mathbf{0}, \ \Sigma(\Lambda))}\left(\left[C_3 \ + \ C_6 \left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]\right) \tag{2.35}$$

where $\mathcal{N}_p(\mathbf{0}, \ \Sigma(\Lambda))$ stands for multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\Sigma(\Lambda)$ (defined in (2.33)). Observe that

$$E_{\mathcal{N}_p(\mathbf{0}, \ \Sigma(\Lambda))}\left(\left[C_3 \ + \ C_6 \left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]\right)$$
$$= \ C_3 \ + \ C_6 \ E_{\mathcal{N}_p(\mathbf{0}, \ \Sigma(\Lambda))}\left[\left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T \Omega(\Lambda)^{1/2} \Sigma(\Lambda) \Omega(\Lambda)^{1/2} (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]$$
$$\leq \ C_3 \ + \ C_6 \ \mathrm{eig}_{\max}^{n\nu/2}(\Sigma(\Lambda)) \ E_{\mathcal{N}_p(\mathbf{0}, \ \Sigma(\Lambda))}\left[\left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T \Omega(\Lambda) (\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]$$

$$
\begin{aligned}
&= C_3 \;+\; C_6 \;\mathrm{eig}_{\max}^{n\nu/2}\left(\Sigma\left(\Lambda\right)\right) \\
&\qquad\qquad E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} + 2\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} + \tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}}\right],
\end{aligned} \tag{2.36}
$$

since $\Sigma\left(\Lambda\right) \;\preceq\; \mathrm{eig}_{\max}\left(\Sigma\left(\Lambda\right)\right) I_p$, and

$$
\begin{aligned}
& E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} + 2\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} + \tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}}\right] \\
&\le\; E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} + 2\left|\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right| + \tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}}\right] \\
&\le\; 3^{n\nu/2}\; E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right)^{\frac{n\nu}{2}} + \left(2\left|\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right|\right)^{\frac{n\nu}{2}}\right. \\
&\qquad\qquad\qquad\qquad \left. +\; \left(\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}}\right] \\
&\left[\text{since for any } a,b,c,n \in \mathbb{R}_+ \cup \{0\},\; (a+b+c)^n \le 3^n\left(a^n + b^n + c^n\right)\right] \\
&=\; 3^{n\nu/2}\left[E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right)^{\frac{n\nu}{2}}\right] + E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(2\left|\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right|\right)^{\frac{n\nu}{2}}\right]\right. \\
&\qquad\qquad\qquad \left. +\, E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}}\right]\right].
\end{aligned} \tag{2.37}
$$

We will show that each term in (2.37) is uniformly bounded in $\boldsymbol{\lambda}$. Note that if $\boldsymbol{\theta} \;\sim\; \mathcal{N}_p\left(\mathbf{0},\;\Sigma\left(\Lambda\right)\right)$, then $\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} \;\sim\; \chi^2_p$. Hence $E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right)^{\frac{n\nu}{2}}\right]$ is a finite quantity free of $\boldsymbol{\lambda}$.

Again, using the fact that $2\left|\mathbf{a}^T\mathbf{b}\right| \le \mathbf{a}^T\mathbf{a} + \mathbf{b}^T\mathbf{b}$, and taking $\mathbf{a} = \Omega\left(\Lambda\right)^{\frac{1}{2}}\tilde{\mathbf{c}}$ and $\mathbf{b} = \Omega\left(\Lambda\right)^{\frac{1}{2}}\boldsymbol{\theta}$, we get

$$
\begin{aligned}
2\left|\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right| \;&\le\; \tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}} \;+\; \boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} \\
&\le\; \mathrm{eig}_{\max}\left(\Omega\left(\Lambda\right)\right)\; \tilde{\mathbf{c}}^T\tilde{\mathbf{c}} \;+\; \boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta} \\
&\left[\text{since } \Omega\left(\Lambda\right) \;\preceq\; \mathrm{eig}_{\max}\left(\Omega\left(\Lambda\right)\right) I_p\right].
\end{aligned}
$$

However, from the expression of $\Omega\left(\Lambda\right)$ in (2.32), it is easy to see that $\Omega\left(\Lambda\right) \preceq I_p$ and hence $\mathrm{eig}_{\max}\left(\Omega\left(\Lambda\right)\right) \le 1$. It follows that

$$
\begin{aligned}
& E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(2\left|\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right|\right)^{\frac{n\nu}{2}}\right] \\
&\le\; 2^{n\nu/2}\; E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\tilde{\mathbf{c}}^T\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}} + \left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right)^{\frac{n\nu}{2}}\right] \\
&=\; 2^{n\nu/2}\;\left[\left(\tilde{\mathbf{c}}^T\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}} + E_{\mathcal{N}_p(\mathbf{0},\;\Sigma(\Lambda))}\left[\left(\boldsymbol{\theta}^T\Omega\left(\Lambda\right)\boldsymbol{\theta}\right)^{\frac{n\nu}{2}}\right]\right] \\
&=\; C_9 \quad (\text{say}),
\end{aligned}
$$

where C_9 is finite and independent of $\boldsymbol{\lambda}$ based on the observations above. Finally, since all eigenvalues of $\Omega\left(\Lambda\right)$ are non-negative and bounded above by 1, it follows that

$$
\left(\tilde{\mathbf{c}}^T\Omega\left(\Lambda\right)\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}} \le \left(\tilde{\mathbf{c}}^T\tilde{\mathbf{c}}\right)^{\frac{n\nu}{2}}.
$$

Using (2.36) and (2.37), we get

$$
\begin{aligned}
& E_{\mathcal{N}_p(\mathbf{0},\ \Sigma(\Lambda))}\left(\left[C_3 \ + \ C_6\left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T(\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]\right) \\
& \leq \ C_3 \ + \ C_{12} \ \mathrm{eig}_{\max}^{n\nu/2}(\Sigma(\Lambda)) \\
& = \ C_3 \ + \ C_{12}\left(2d_{\max}^2+1\right)^{\frac{n\nu}{2}} \qquad \text{(by the expression of } \Sigma(\Lambda) \text{ in (2.33))},
\end{aligned}
$$

where C_{12} is an appropriate constant independent of $\boldsymbol{\lambda}$ and $d_{\max}$ is the largest element of the diagonal matrix D in the expression of (2.33). Plugging this in (2.35), we get the following upper bound for the inner integral in (2.34).

$$
\begin{aligned}
& \int_{\mathbb{R}^p}\left[C_3 \ + \ C_6\left((\boldsymbol{\theta}+\tilde{\mathbf{c}})^T(\boldsymbol{\theta}+\tilde{\mathbf{c}})\right)^{\frac{n\nu}{2}}\right]\exp\left[-\frac{1}{2}\,\boldsymbol{\theta}^T\Omega(\Lambda)\,\boldsymbol{\theta}\right]\,d\boldsymbol{\theta} \\
& \leq \ (2\pi)^{p/2}\sqrt{\det(\Sigma(\Lambda))}\ \left[C_3 \ + \ C_{12}\left(2d_{\max}^2+1\right)^{\frac{n\nu}{2}}\right].
\end{aligned}
$$

This in turn leads to the following upper bound for the integral I of interest.

$$
\begin{aligned}
I \ \leq \ C_7\,(2\pi)^{p/2}\int_{\mathbb{R}_+^n} & \max\left\{1,\lambda_{\max}^{p/2}\right\}\times\left[\prod_{i=1}^n\lambda_i^{\frac{\nu}{2}-1}\right]\times\ \exp\left[-\frac{\nu}{2}\sum_{i=1}^n\lambda_i\right] \\
& \times\sqrt{\det(\Sigma(\Lambda))}\ \left[C_3 \ + \ C_{12}\left(2d_{\max}^2+1\right)^{\frac{n\nu}{2}}\right]\ d\boldsymbol{\lambda}. \qquad (2.38)
\end{aligned}
$$

Our final steps will be to bound the (single) integral in (2.38) separately in the low and high-dimensional settings.

Step VII: An upper bound for the integral in (2.38) in the low- dimensional setting. Suppose $n \geq p$. By the expression of $\Sigma(\Lambda)$ in (2.33) for $n \geq p$, we have

$$
\begin{aligned}
\det(\Sigma(\Lambda)) &= \det\left(V\left(2D^2+I_p\right)V^T\right) \\
&= \prod_{i=1}^p\left(2d_i^2+1\right) \\
&\leq \left(2d_{\max}^2+1\right)^p, \qquad (2.39)
\end{aligned}
$$

where $\{d_i\}_{i=1}^p$ are the singular values of A, and $d_{\max}$ is the largest singular value of $A = \Lambda^{1/2}W$ in (2.29). Note that

$$
d_{\max}^2 \ = \ \mathrm{eig}_{\max}\left(AA^T\right) = \ \mathrm{eig}_{\max}\left(\Lambda^{1/2}WW^T\Lambda^{1/2}\right). \qquad (2.40)
$$

Since WW^T is a fixed positive semi-definite matrix, there exists a large positive real number C_{13} such that

$$
\Lambda^{1/2}WW^T\Lambda^{1/2} \preceq C_{13}\Lambda \preceq C_{13}\lambda_{\max}I_n,
$$

where $\lambda_{\max}$ is the largest element in the matrix Λ. It follows from (2.40) that $d^2_{\max} \leq C_{13}\lambda_{\max}$. Using (2.38), (2.39) and the c_r-inequality, we get

$$\begin{aligned} I \;\leq\; & C_7\,(2\pi)^{p/2} \int_{\mathbb{R}^n_+} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] \\ & \qquad \times \left(2d^2_{\max}+1\right)^{\frac{p}{2}} \left[C_3 \;+\; C_{12}\left(2d^2_{\max}+1\right)^{\frac{n\nu}{2}}\right] d\boldsymbol{\lambda} \\ \leq\; & C_7\,(2\pi)^{p/2} \int_{\mathbb{R}^n_+} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] \\ & \qquad \times (2C_{13}\lambda_{\max}+1)^{\frac{p}{2}} \left[C_3 \;+\; C_{12}\,(2C_{13}\lambda_{\max}+1)^{\frac{n\nu}{2}}\right] d\boldsymbol{\lambda} \\ \leq\; & C_7\,(2\pi)^{p/2} \int_{\mathbb{R}^n_+} \left(1+\lambda_{\max}^{p/2}\right) \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] \\ & \qquad \times 2^{p/2}\left(C_{14}\lambda_{\max}^{p/2}+1\right) \left[C_3 \;+\; C_{12}2^{n\nu/2}\left(C'_{14}\lambda_{\max}^{n\nu/2}+1\right)\right] d\boldsymbol{\lambda}, \end{aligned}$$

where $C_{14} = (2C_{13})^{p/2}$ and $C'_{14} = (2C_{13})^{n\nu/2}$. Expanding the product of the polynomial terms in $\lambda_{\max}$ in the integrand gives

$$\begin{aligned} I \leq\; & C_{15} \int_{\mathbb{R}^n_+} \left(\lambda_{\max}^{p+\frac{n\nu}{2}} + \lambda_{\max}^{p} + \lambda_{\max}^{(p+n\nu)/2} + \lambda_{\max}^{p/2} + \lambda_{\max}^{n\nu/2} + 1\right) \\ & \qquad \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] d\boldsymbol{\lambda}, \end{aligned}$$

for an appropriate finite constant C_{15} which does not depend on $\boldsymbol{\lambda}$. Using the fact that $\lambda_{\max}^r \leq \sum_{j=1}^{n} \lambda_j^r$ for any positive r, we get

$$\begin{aligned} I \leq\; & C_{15} \sum_{j=1}^{n} \int_{\mathbb{R}^n_+} \left(\lambda_j^{p+\frac{n\nu}{2}} + \lambda_j^{p} + \lambda_j^{(p+n\nu)/2} + \lambda_j^{p/2} + \lambda_j^{n\nu/2} + 1\right) \\ & \qquad \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times\; \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] d\boldsymbol{\lambda} \\ \leq\; & C_{15} \sum_{j=1}^{n} \left(\int_{\mathbb{R}_+} \left(\lambda_j^{p+\frac{n\nu}{2}} + \lambda_j^{p} + \lambda_j^{(p+n\nu)/2} + \lambda_j^{p/2} + \lambda_j^{n\nu/2} + 1\right) \lambda_j^{\frac{\nu}{2}-1} \right. \\ & \qquad \left. \times \exp\left[-\frac{\nu}{2}\lambda_j\right] d\lambda_j\right) \\ & \qquad \times \left(\prod_{i=1, i\neq j}^{n} \int_{\mathbb{R}_+} \lambda_i^{\frac{\nu}{2}-1}\; \exp\left[-\frac{\nu}{2}\lambda_i\right]\; d\lambda_i\right). \end{aligned} \tag{2.41}$$

Since all of the terms in the above bound are integrals over unnormalized gamma densities (with strictly positive and finite shape and rate parameters), it follows that $I < \infty$.

Step VIII: An upper bound for the integral in (2.38) in the high-dimensional setting. Suppose $n < p$. By the expression of $\Sigma(\Lambda)$ in (2.33) for $n < p$, we have

$$\begin{aligned}
\det(\Sigma(\Lambda)) &= \det\left(U\left(2D^2 + I_n\right)U^T + U_0 U_0^T\right) \\
&= \det\left(\begin{bmatrix} \left(2D^2 + I_n\right) & 0 \\ 0 & I_{p-n} \end{bmatrix}\right) \\
&= \prod_{i=1}^{n} \left(2d_i^2 + 1\right) \\
&\leq \left(2d_{\max}^2 + 1\right)^n .
\end{aligned} \tag{2.42}$$

By a similar argument as for the $n \geq p$ setting, we get $d_{\max}^2 \leq C'_{13}\lambda_{\max}$ for an appropriate constant C'_{13} not depending on $\boldsymbol{\lambda}$. Using (2.38), (2.42) and the c_r-inequality, we get

$$\begin{aligned}
I &\leq C_7 (2\pi)^{p/2} \int_{\mathbb{R}_+^n} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] \\
&\qquad \times \left(2d_{\max}^2 + 1\right)^{\frac{n}{2}} \left[C_3 + C_{12}\left(2d_{\max}^2 + 1\right)^{\frac{n\nu}{2}}\right] d\boldsymbol{\lambda} \\
&\leq C_7 (2\pi)^{p/2} \int_{\mathbb{R}_+^n} \max\left\{1, \lambda_{\max}^{p/2}\right\} \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] \\
&\qquad \times \left(2C'_{13}\lambda_{\max} + 1\right)^{\frac{n}{2}} \left[C_3 + C_{12}\left(2C'_{13}\lambda_{\max} + 1\right)^{\frac{n\nu}{2}}\right] d\boldsymbol{\lambda} \\
&\leq C_7 (2\pi)^{p/2} \int_{\mathbb{R}_+^n} \left(1 + \lambda_{\max}^{p/2}\right) \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] \\
&\qquad \times 2^{n/2}\left(C_{13}^*\lambda_{\max}^{n/2} + 1\right) \\
&\qquad \times \left[C_3 + C_{12}2^{n\nu/2}\left(C_{14}^*\lambda_{\max}^{n\nu/2} + 1\right)\right] d\boldsymbol{\lambda},
\end{aligned}$$

where $C_{13}^* = (2C'_{13})^{n/2}$ and $C_{14}^* = (2C'_{13})^{n\nu/2}$ similar to the low-dimensional setting. Expanding the product of the polynomial terms in $\lambda_{\max}$ in the integrand gives

$$\begin{aligned}
I \leq C'_{15} \int_{\mathbb{R}_+^n} \Bigg(& \lambda_{\max}^{\frac{p+n+n\nu}{2}} + \lambda_{\max}^{\frac{n+n\nu}{2}} + \lambda_{\max}^{\frac{p+n}{2}} + \lambda_{\max}^{\frac{n}{2}} \\
& + \lambda_{\max}^{(p+n\nu)/2} + \lambda_{\max}^{n\nu/2} + \lambda_{\max}^{p/2} + 1 \Bigg) \\
& \times \left[\prod_{i=1}^{n} \lambda_i^{\frac{\nu}{2}-1}\right] \times \exp\left[-\frac{\nu}{2}\sum_{i=1}^{n}\lambda_i\right] d\boldsymbol{\lambda},
\end{aligned}$$

for an appropriate finite constant C'_{15} which does not depend on $\boldsymbol{\lambda}$. Using the fact that $\lambda_{\max}^r \leq \sum_{j=1}^{n} \lambda_j^r$ for any positive r, and similar arguments regarding gamma integrals as in (2.41) for the low-dimensional $n \geq p$ setting, it follows that $I < \infty$ in the high-dimensional $n < p$ setting as well. This establishes the trace-class property of the DA Markov chain. □

As discussed in the introduction, the trace-class property established above implies compactness of the Markov operator K, which implies that the corresponding DA Markov chain is geometrically ergodic.

Corollary 1. *The DA Markov chain with transition density k (in (2.1)) is geometrically ergodic for an arbitrary choice of the design matrix X, sample size n, number of predictors p, degrees of freedom $\nu > 0$, prior mean vector $\boldsymbol{\beta}_a$, and (positive definite) prior precision matrix Σ_a.*

Remark 2.1. We note that the arguments posited in Step IV of the proof above directly utilize properties of a t_ν-cdf embodying the (inverse) robit link, and cannot be directly extended to handle a probit model. In particular, equation (2.26) produces a Mill's ratio type upper bound for the product term involving reciprocal cumulative t_ν-probabilities. This upper bound is an $n\nu$-degree multi-linear polynomial in the elements of $\boldsymbol{\theta}$, the linearly transformed $\boldsymbol{\beta}$ variables. The polynomial nature of this upper bound is crucial in ensuring soundness of the subsequent arguments; e.g., this polynomial upper bound in steps V and VI combined produces an upper bound for the inner $\boldsymbol{\theta}$ integral in (2.16) in terms of moments of a certain multivariate normal distribution. Heuristically, to extend the presented proof to the case of a probit model, one would need to let $\nu \to \infty$, followed by setting $\lambda \equiv 1$ everywhere in the integral (2.12) and removing all λ-only terms and the outer λ integral in (2.11). However for non-finite ν the $n\nu$-degree polynomial upper bound (2.26) is invalid and instead one obtains an upper bound that is exponential in $\boldsymbol{\theta}^T\boldsymbol{\theta}$ [see 2]. This extra exponential term renders several subsequent arguments posited in our proof inapplicable. Consequently, despite lacking any λ terms or an outer λ integral, the target trace class integral for a probit model Markov chain cannot be proven to be finite via a simple adaptation of our arguments. To prove the trace class property of the Bayesian probit model Markov chain separate arguments for the extra exponential term arising out of the Mill's ratio upper bound are made in [2], and regularity conditions bounding the eigenvalues of a product matrix involving the design matrix X and the prior precision matrix Σ_a are imposed. Note that, no such regularity condition is needed for the robit model in Theorem 1.

3. Numerical illustrations

This section presents numerical illustrations to compare/contrast the convergence properties of the robit DA and some other relevant Markov chains. To examine both the low and high-dimensional settings, we consider two real data sets, viz., the Lupus data ($n > p$) from [18] and the prostate cancer data ($n < p$) from [4]. We note at the outset that with a view to the main goal/contribution of this paper, viz., theoretical convergence analysis for (robit) Markov chains, our focus in this section centers entirely around exemplification of the said convergences. Prior elicitation for statistical inference are beyond the scope of this section and the paper; we consider the data-driven Zellner's g-prior following [2] in the first example, and independent standard normal priors in the second example to run the respective Markov chains.

As noted in the Introduction, a trace-class property ensures guaranteed improvements in the convergence of a DA algorithm by *sandwiching* (see Corollary 2 below). To exemplify/visualize this improvement in the current context we alongside consider a sandwich algorithm obtained by inserting an inexpensive random generation step in between the two steps of Algorithm 1. More specifically, we consider the sandwich algorithm from [16] which inserts a univariate random gamma generation as presented in Algorithm 2. Note that both the original DA algorithm and its sandwiched version share the same target stationary distribution for $\boldsymbol{\beta}$.

Algorithm 2: $(m+1)$-st Iteration of the Robit Sandwich Algorithm

1. Make independent draws from the $Tt_\nu(\mathbf{x}_i^T\boldsymbol{\beta}^{(m)}, y_i)$ distributions for $1 \le i \le n$. Denote the respective draws by $z_1, z_2, \cdots, z_n$. Draw λ_i from the Gamma$\left(\frac{\nu+1}{2}, \frac{\nu+\left(z_i-\mathbf{x}_i^T\boldsymbol{\beta}^{(m)}\right)^2}{2}\right)$ distribution.
2. Generate $h^2 \sim$ Gamma$\left(\frac{n}{2}, \frac{\mathbf{z}^T\Lambda^{1/2}(I-Q)\Lambda^{1/2}\mathbf{z}}{2}\right)$, where $Q = \Lambda^{1/2}X\left(X^T\Lambda X + \Sigma_a\right)^{-1}X^T\Lambda^{1/2}$, and subsequently define $z_i' = hz_i$; $1 \le i \le n$.
3. Draw $\boldsymbol{\beta}^{(m+1)}$ from the $\mathcal{N}_p\left((X^T\Lambda X + \Sigma_a)^{-1}(X^T\Lambda\mathbf{z}' + \Sigma_a\boldsymbol{\beta}_a),\ (X^T\Lambda X + \Sigma_a)^{-1}\right)$ distribution.

The trace class property of the robit DA chain (Theorem 1) along with results in [8] imply that the following properties hold for the sandwich chain.

Corollary 2. *The sandwich Markov chain described in Algorithm 2 is trace class (and hence geometrically ergodic) for an arbitrary choice of the design matrix X, sample size n, number of predictors p, degrees of freedom $\nu > 0$, prior mean vector $\boldsymbol{\beta}_a$, and (positive definite) prior precision matrix Σ_a. Furthermore, if $(\lambda_i)_{i=0}^\infty$ and $(\lambda_i^*)_{i=0}^\infty$ denote the non-increasing sequences of eigenvalues corresponding to the robit DA and sandwich operators respectively, then $\lambda_i^* \le \lambda_i$ for every $i \ge 0$, with at least one strict inequality.*

To facilitate comparison in each example below we consider three robit models: (i) one with degree of freedom $\nu = 1$ (the *cauchit* model), (ii) one with $\nu = 3$, and (iii) one with $\nu = 1000$. From an application point of view, both (i) and (ii) are expected to produce robust binary models, with (i) inducing more robustness than (ii). By contrast, (iii) is expected to mimic the conventional probit model, owing to the limiting property of a t distribution with large degrees of freedom. For each model we consider two Markov chains – the original DA chain and a corresponding sandwich chain.

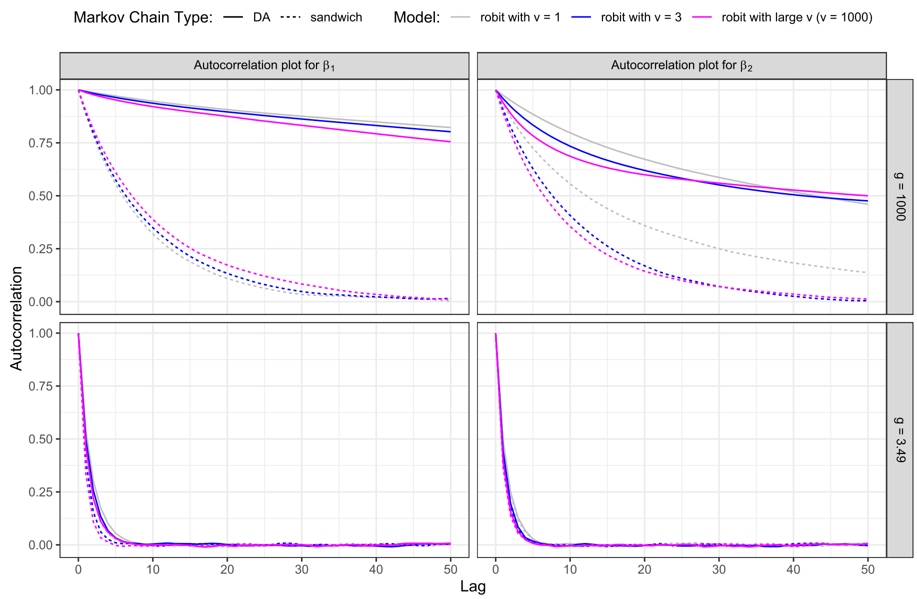

Fig 1. *Autocorrelation plots for Markov chains run on the Lupus data set.*

3.1. Low dimensional ($n > p$) setting: Lupus data set

The Lupus data set of [18] comprises observations on two antibody molecule predictor variables and a binary outcome variable cataloging occurrences of latent membranous lupus nephritis among $n = 55$ patients. Interest lies in regressing the binary outcome on the predictors; in this regression we include an intercept term which effectively makes the number of predictors to be $p = 3$. The data set has been previously considered in the context of convergence analyses of the probit model DA and sandwich Markov chains [18, 17, 2]. Here we use it to illustrate convergences of the robit model Markov chains. Following [2] we assign on the regression coefficient vector $\boldsymbol{\beta}$ the Zellner's g-prior $\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0},\ g(X^T X)^{-1})$ with two choices of g: (a) $g = 1000$ which induces a diffuse prior on $\boldsymbol{\beta}$, and (b) $g = 3.49$ which ensures the trace-class property of the probit DA algorithm [2]. For this data set we thus collectively consider 12 Markov chains from 3 models, 2 priors, and 2 Markov chain types (DA or sandwich). All 12 chains are initiated at the maximum likelihood estimates $\beta_0 = -1.778$ (intercept), $\beta_1 = 4.374$, and $\beta_2 = 2.428$ obtained from the probit model. We run each chain for 100,000 iterations, after discarding the initial 10,000 iterations as burn-in; the adequacy of the burn-in period is justified through traceplots (Appendix Figures D.1 - D.3). We subsequently use the retained realizations from all the 12 Markov chains to compute (a) Markov chain autocorrelations up to lag 50, and (b) running means for each of the two non-intercept regression coefficients β_1 and β_2.

The autocorrelations and running means are displayed in Figures 1 and 2 as plot-matrices with the rows corresponding to priors (g priors with $g = 1000$ and

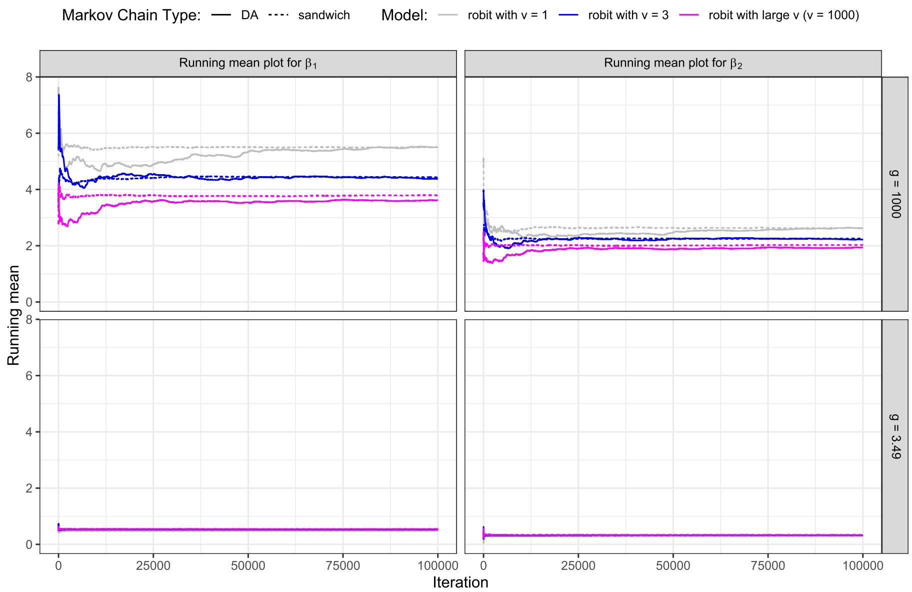

FIG 2. *Running mean plots for the Markov chains run on the Lupus data set.*

$g = 3.49$) and columns corresponding to $\boldsymbol{\beta}$ components. Individual plots in the plot-matrices display as line diagrams autocorrelations (y-axis) plotted against lags (x-axis) in Figure 1, or running means (y-axis) plotted against Markov chain iterations (x-axis) in Figure 2. A separate line is drawn for each model/Markov chain type combination (6 lines in total in each plot). The lines are color coded by models with the probit model, the robit model with small ν, and the robit model with large ν being displayed as red, gray, and blue lines respectively. On the other hand, Markov chain types are displayed via line types: solid and dashed lines are used for DA and sandwich chains respectively.

The following three observations are made from these two plots. First, for all three models sandwiching appears to aid substantial improvements in convergence and mixing over the original DA algorithm when the underlying prior is vague ($g = 1000$). This is demonstrated by both lowered autocorrelations in Figure 1 (top rows/panels) and stabler running means in Figure 2 (top rows/panels) for the $\boldsymbol{\beta}$-components in the sandwich chains. Interestingly, when viewed as functions of ν, the improvement pattern appears to differ in β_1 and β_2. For β_2 (top-right panel of Figure 1) the $\nu = 1000$ sandwich chain appears to enjoy the least autocorrelation, followed by the $\nu = 3$, and $\nu = 1$ sandwich chains, with the third having noticeably higher autocorrelations than the former two. In contrast, a less noticeable, but opposite pattern is observed for β_1 (top-left corner of Figure). The improvements in sandwich algorithms are not noticeable when the more informative prior with $g = 3.49$ is used. There, the DA and the sandwich chains display similar convergence properties, and the lines from the different model/Markov chain type combinations all effectively get superimposed at the displayed scale (bottom rows/panels in Figures 1 and 2). Second, the running means from the DA chain are less stable than the sandwich chain; however, they do converge to the same limit *within the same model*. This is clearly expected since both the DA and sandwich chain have the same stationary distribution for $\boldsymbol{\beta}$. This is particularly well-documented in the upper panels of Figure 2. Third, when the vague prior ($g = 1000$) is used, the running means from robit models with different ν's owing to the differences in the likelihood caused by the ν-values. However, when the more informative ($g = 3.45$) prior is used the posteriors become less impacted by the systematic differences in the likelihoods and are more driven by the prior information; a fact well visualized in the lower panels of Figure 2. There, all chains from all models appear to share the same *limiting* running means at the scale displayed.

3.2. High dimensional ($n < p$) setting: Prostate data set

In the second example we consider the prostate cancer dataset from [4]. The dataset records gene expressions of 50 normal and 52 prostate tumor samples at 6033 arrays, of which we select the first 150 arrays for our analysis. We are interested in regressing the binary cancer status (normal = 0, tumor = 1) on these selected 150 expression arrays (predictors). Similar to the analysis done in the previous section, we include an intercept term to the regression model to obtain the number of predictors $p = 151$ which is bigger than the total sample size of $n = 102$. We consider three robit models as before: (a) with $\nu = 1$ (cauchit model), (b) with $\nu = 3$, and (c) with (large) $\nu = 1000$, and in each model assign *independent standard normal priors* on the components of the regression coefficient vector $\boldsymbol{\beta}$. For each model we then run two Markov chains – the original DA chain, and the corresponding sandwich chain. All 6 chains are initiated at $\boldsymbol{\beta} = \mathbf{0}$ and are run for 100,000 iterations, *after* discarding the first 10,000 iterations as burn-in. Subsequently, the (un-normalized) log-likelihood $\mathrm{lik}(\boldsymbol{\beta})$ and (un-normalized) log-posterior density $\mathrm{lpd}(\boldsymbol{\beta})$ values are calculated as univariate functions of $\boldsymbol{\beta}$ on the retained realizations of each Markov chain. Here

$$\mathrm{lik}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ y_i \log F(\boldsymbol{x}_i^T \boldsymbol{\beta}) + (1 - y_i) \log[1 - F(\boldsymbol{x}_i^T \boldsymbol{\beta})] \right\}, \text{ and}$$
$$\mathrm{lpd}(\boldsymbol{\beta}) = \mathrm{lik}(\boldsymbol{\beta}) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\beta},$$

and F is the normal/t CDF associated with the probit/robit model. Finally for these computed log-likelihoods and log-posterior densities we calculate the Markov chain autocorrelations upto lag 50 and running means as done in the previous Lupus example. The resulting values are displayed in Figures 3 and 4 respectively. These figures follow the same color and line type conventions as considered in Figures 1 and 2.

The following observations are made from these figures. First, as expected, the sandwich chains (broken lines) are observed to have (moderately) better con-

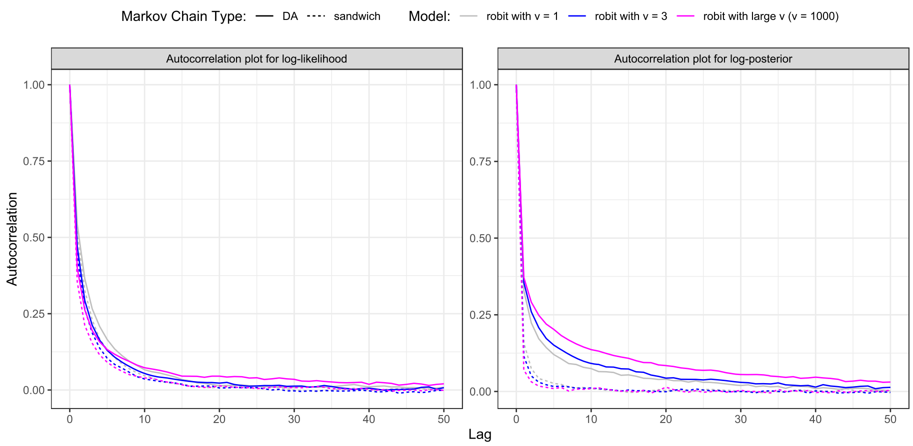

FIG 3. *Autocorrelation plots for the Markov chains run on the Prostate data set.*

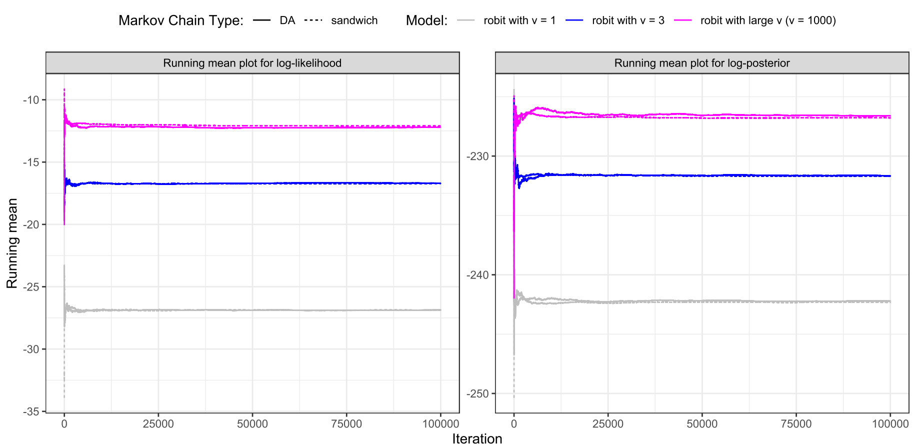

FIG 4. *Running mean plots for the Markov chains run on the Prostate data set.*

vergences, i.e., smaller auto-correlations (Figure 3) and stabler running means (Figure 4) than the original DA chains (solid lines). The improvement is more prominent among the log-posterior density values (right panel in each figure). Second, there appears to be some small/moderate difference in the autocorrelations for the DA chains corresponding to different ν values (noticeable more on the right, log-posterior panel on Figure 4). The cauchit $\nu = 1$ DA chain appears to have the smallest autocorrelations, and the autocorrelations seem to marginally increase with ν. By contrast, all three sandwich chains appear to have similar auto-correlations. Third, for both the DA and sandwich chains, the "limiting" running means for the log-likelihood and log-posterior density values from the three robit models differ systematically owing to the difference in the model likelihoods induced by differences in ν.

Appendix A: A Mill's ratio type result for student's t distribution

Lemma A.1. *For $t > 0$ we have*

$$\frac{1}{(1 - F_\nu(t))} \leq \frac{\left(t^2 + \nu\right)^{\frac{\nu}{2}}}{\kappa},$$

where $\kappa = \Gamma\left((\nu+1)/2\right)\nu^{\nu/2-1}/\left(\sqrt{\pi}\Gamma\left(\nu/2\right)\right)$, and $F_\nu(.)$ is the cdf of $t_\nu(0,1)$.

Proof. Firstly, let us introduce the incomplete beta function ratio, defined by

$$I_p(a,b) = \frac{1}{B(a,b)} \int_0^p u^{a-1}(1-u)^{b-1} du, \quad 0 < p < 1, \tag{A.1}$$

where, $a, b \in \mathbb{R}_+$, and $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Let, $f_\nu(.)$ be the pdf of $t_\nu(0,1)$. The cumulative distribution function $F_\nu(.)$ can be written in terms of I, the incomplete beta function ratio. From Chapter 28 of [12], we have for $t > 0$,

$$F_\nu(t) = \int_{-\infty}^t f_\nu(u) du = 1 - \frac{1}{2} I_{x(t)}\left(\frac{\nu}{2}, \frac{1}{2}\right), \tag{A.2}$$

where

$$x(t) = \frac{\nu}{t^2 + \nu}.$$

Now, from the definition (A.1) of the incomplete beta function ratio, it follows that

$$\begin{aligned} I_{x(t)}\left(\frac{\nu}{2}, \frac{1}{2}\right) &= \frac{1}{B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_0^{x(t)} u^{\frac{\nu}{2}-1}(1-u)^{\frac{1}{2}-1} du \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^{x(t)} u^{\frac{\nu}{2}-1}(1-u)^{\frac{1}{2}-1} du \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^{\frac{\nu}{t^2+\nu}} \frac{u^{\frac{\nu}{2}-1}}{\sqrt{1-u}} du. \end{aligned} \tag{A.3}$$

For the integral in the right hand side of (A.3), we have

$$\begin{aligned} \int_0^{\frac{\nu}{t^2+\nu}} \frac{u^{\frac{\nu}{2}-1}}{\sqrt{1-u}} du &\geq \int_0^{\frac{\nu}{t^2+\nu}} u^{\frac{\nu}{2}-1} \quad \left[\text{since } \frac{1}{\sqrt{1-u}} > 1 \text{ for } u \in \left(0, \frac{\nu}{t^2+\nu}\right)\right] \\ &= \frac{2}{\nu}\left[u^{\frac{\nu}{2}}\right]_{u=0}^{u=\frac{\nu}{t^2+\nu}} \\ &= \frac{2\ \nu^{\nu/2-1}}{(t^2+\nu)^{\frac{\nu}{2}}}. \end{aligned} \tag{A.4}$$

Using (A.2), (A.3) and (A.4), we finally have

$$I_{x(t)}\left(\frac{\nu}{2}, \frac{1}{2}\right) \geq \frac{2\ \Gamma\left(\frac{\nu+1}{2}\right)\ \nu^{\nu/2-1}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \times \frac{1}{(t^2+\nu)^{\frac{\nu}{2}}}$$

$$\Longrightarrow\ 1-F_\nu(t)\ \geq\ \frac{\Gamma\left(\frac{\nu+1}{2}\right)\ \nu^{\nu/2-1}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\times\frac{1}{\left(t^2+\nu\right)^{\frac{\nu}{2}}}$$
$$\Longrightarrow\ \frac{1}{\left(1-F_\nu\left(t\right)\right)}\leq\frac{\left(t^2+\nu\right)^{\frac{\nu}{2}}}{\kappa},$$

where $\kappa=\Gamma\left(\left(\nu+1\right)/2\right)\nu^{\nu/2-1}/\left(\sqrt{\pi}\Gamma\left(\nu/2\right)\right)$. □

Appendix B: Geometric ergodicity of the robit DA chain using drift and minorization in the $n\geq p$ setting

In this section, we present a tighter drift and minorization analysis, and establish geometric ergodicity of the robit DA chain under significantly weaker conditions that those in [16]. We consider the $n\geq p$ setting here, and consider the high-dimensional $n<p$ setting in Appendix C.

Let W be an $n\times p$ matrix whose ith row is w_i^T where $w_i=x_iI_{\{0\}}\left(y_i\right)-x_iI_{\{1\}}\left(y_i\right)$. Let us define the following two notations which have been derived in the proof of the theorem below.

$$\rho_1=\rho'\left(\frac{\nu+1}{\nu a}\right)\qquad\text{and}\qquad\rho_2\ =\ \rho'\left(\frac{\nu+1}{\nu a}\right)\left(\sqrt{\frac{\left(\nu+3\right)^2}{\left(\nu+1\right)\left(\nu+3\right)\ +\ a^2\nu^2}}\right),$$

where $\rho'\leq 1$ is as defined in Appendix B in [16]. In fact, it is shown in [16] that $\rho'<1$ if X has full column rank. By establishing a drift condition for the DA algorithm we prove the following theorem.

Theorem B.1. *The DA algorithm is geometrically ergodic if*

(A) $\nu>2$,
(B) *The design matrix X has full column rank, and* $\Sigma_a=aX^TX$,
(C) $min\{\rho_1,\rho_2\}\ <\ 1$.

Proof. One approach to prove the geometric convergence of the robit DA Markov chain is to prove the existence of a *drift function* $V:\mathbb{R}^p\to\mathbb{R}_+$ which satisfies the following two conditions:

1. $\int_{\mathbb{R}^p}V(\boldsymbol{\beta})k\left(\boldsymbol{\beta}_0,\boldsymbol{\beta}\right)d\boldsymbol{\beta}\ \leq\ \rho V(\boldsymbol{\beta}_0)+L$, for some $\rho\in[0,1)$ and $L>0$.
2. If $V(\boldsymbol{\beta})\leq\alpha$ for $\alpha>0$, then $||\boldsymbol{\beta}||_2\leq d$ for some d. Such function V is said to be an *unbounded off compact sets.*

As in [16], we will use $V(\boldsymbol{\beta})=\boldsymbol{\beta}^TX^TX\boldsymbol{\beta}$ to establish the above two conditions. It is easy to see that the second condition is satisfied since $X_{n\times p}$ is assumed to have full column rank p.

We now prove the first condition. Note that the robit DA Markov chain transitions from the state $\boldsymbol{\beta}_0$ to the state $\boldsymbol{\beta}$ through the following two steps: (i) by generating the intermediate latent variables $(\boldsymbol{\lambda},\mathbf{z})$ first and, (ii) then generating a new $\boldsymbol{\beta}$ from the relevant conditional distribution. This can be represented by

$$\boldsymbol{\beta}_0\to(\boldsymbol{\lambda},\mathbf{z})\to\boldsymbol{\beta}.$$

Then, the L.H.S. in the condition (1) can be written as:

$$\begin{aligned}
&\int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k\left(\boldsymbol{\beta}_0, \boldsymbol{\beta}\right) d\boldsymbol{\beta} \\
&= E\left[\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= E\left[E\left[\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \mid \boldsymbol{\lambda}, \mathbf{z}, \boldsymbol{\beta}_0, \mathbf{y}\right] \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= E\left[E\left[\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \mid \boldsymbol{\lambda}, \mathbf{z}, \mathbf{y}\right] \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= E\left[E\left[||X\boldsymbol{\beta}||_2^2 \mid \boldsymbol{\lambda}, \mathbf{z}, \mathbf{y}\right] \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right],
\end{aligned} \tag{B.1}$$

where the inner expectation is w.r.t. the conditional distribution of $\boldsymbol{\beta}$ given $(\boldsymbol{\lambda}, \mathbf{z}, \mathbf{y})$, and the outer expectation is w.r.t. the conditional distribution of $(\boldsymbol{\lambda}, \mathbf{z})$ given $(\boldsymbol{\beta}, \mathbf{y})$.

Since

$$\boldsymbol{\beta}\Big|\boldsymbol{\lambda}, \mathbf{z}, \mathbf{y} \sim N_p\left(\left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right), \left(X^T\Lambda X + aX^TX\right)^{-1}\right),$$

it follows that

$$\begin{aligned}
X\boldsymbol{\beta}\Big|\boldsymbol{\lambda}, \mathbf{z}, \mathbf{y} \;\sim\; N_n\Big(&X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right), \\
&X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Big),
\end{aligned} \tag{B.2}$$

where $\mathbf{c} = \Sigma_a\boldsymbol{\beta}_a = aX^TX\boldsymbol{\beta}_a$, is a fixed vector free of the parameters. Using (B.2), the inner expectation in (B.1) is given by

$$\begin{aligned}
E\left(||X\boldsymbol{\beta}||_2^2\Big|\boldsymbol{\lambda}, \mathbf{z}, \mathbf{y}\right) = &\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right)\right|\right|_2^2 \\
&+ \mathrm{tr}\left(X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\right).
\end{aligned} \tag{B.3}$$

Again, since X is a full column rank matrix, by the Singular Value Decomposition, X can be written as

$$X_{n\times p} = U_{n\times p}D_{p\times p}V_{p\times p}^T,$$

where U is a semi-orthogonal matrix, V is an orthogonal matrix, and D is a diagonal matrix with the positive singular values of X being its diagonal entries. Substituting $X = UDV^T$ in the second term in the R.H.S. of (B.3), we get

$$\begin{aligned}
&\mathrm{tr}\left(X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\right) \\
&= \mathrm{tr}\left(UDV^T\left(VDU^T\Lambda UDV^T + aVDU^TUDV^T\right)^{-1}VDU^T\right) \\
&= \mathrm{tr}\left(UDV^T\left(VD\left(U^T\Lambda U + aU^TU\right)DV^T\right)^{-1}VDU^T\right) \\
&= \mathrm{tr}\left(UDV^T\left(DV^T\right)^{-1}\left(U^T\Lambda U + aU^TU\right)^{-1}(VD)^{-1}VDU^T\right)
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\left(U\left(U^T\Lambda U + aU^TU\right)^{-1}U^T\right) \\
&= \operatorname{tr}\left(U\left(U^T\Lambda U + aI\right)^{-1}U^T\right) \qquad [\text{Since } U^TU = I].
\end{aligned}
$$

For matrices A and B, $A \preceq B$ denotes that the matrix $(B-A)$ is a positive semi-definite or non-negative definite (N.N.D) matrix. Since $(U^T\Lambda U + aI)^{-1} \preceq a^{-1}I$, it follows that

$$
(U^T\Lambda U + aI)^{-1} \prec a^{-1}I \implies U\left(U^T\Lambda U + aI\right)^{-1}U^T \preceq a^{-1}UU^T,
$$

and consequently

$$
\operatorname{tr}\left(U\left(U^T\Lambda U + aI\right)^{-1}U^T\right) \le \operatorname{tr}\left(a^{-1}UU^T\right) = a^{-1}\operatorname{tr}\left(U^TU\right) = a^{-1}p. \tag{B.4}
$$

Now, considering the first-term in the R.H.S. of (B.3), we apply triangle inequality to get

$$
\begin{aligned}
&\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right)\right|\right|_2 \\
&\le \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2 + \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right)\right|\right|_2^2 \\
&\le \left(\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2 + \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2\right)^2 \\
&= \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 + \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2^2 \\
&\qquad + 2\left(X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right)^T\left(X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right) \\
&= \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 + \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2^2 \\
&\qquad + 2\left(\frac{X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}}{d}\right)^T \\
&\qquad\qquad \times\left(d\;X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right) \\
&\quad [\text{where, } d > 0 \text{ is an arbitrary real number}] \\
&\le \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 + \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2^2 \\
&\qquad + 2\left|\left(\frac{X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}}{d}\right)^T\right. \\
&\qquad\qquad \left.\times\left(d\;X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 + \left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2^2 \\
&\qquad + \frac{\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2}{d^2} \\
&\qquad + d^2\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2^2 \\
&\text{[since } 2|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||_2^2 + ||\mathbf{y}||_2^2, \text{ for any two vectors } \mathbf{x}, \mathbf{y}] \\
&= \left(1 + \frac{1}{d^2}\right)\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 \\
&\qquad + \left(1 + d^2\right)\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\mathbf{c}\right|\right|_2^2 \\
&\leq \left(1 + \frac{1}{d^2}\right)\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 \\
&\qquad + \left(1 + d^2\right)||X||_2^2\left|\left|\left(X^T\Lambda X + aX^TX\right)^{-1}\right|\right|_2^2||\mathbf{c}||_2^2. \qquad \text{(B.5)}
\end{aligned}
$$

Now by SVD, $X = UDV^T$

$$
\begin{aligned}
\left|\left|\left(X^T\Lambda X + aX^TX\right)^{-1}\right|\right|_2 &= \left|\left|\left(VDU^T\Lambda UDV^T + aVDDV^T\right)^{-1}\right|\right|_2 \\
&= \left|\left|\left[VD\left(U^T\Lambda U + aI\right)DV^T\right]^{-1}\right|\right|_2 \\
&= \left|\left|\left(DV^T\right)^{-1}\left(U^T\Lambda U + aI\right)^{-1}(VD)^{-1}\right|\right|_2 \\
&\leq \left|\left|\left(DV^T\right)^{-1}\right|\right|_2\left|\left|\left(U^T\Lambda U + aI\right)^{-1}\right|\right|_2 \\
&\qquad \times \left|\left|(VD)^{-1}\right|\right|_2 \\
&\leq \frac{1}{a}\left|\left|\left(DV^T\right)^{-1}\right|\right|_2\left|\left|(VD)^{-1}\right|\right|_2 := K_2,
\end{aligned}
$$

where $K_2 > 0$ is a constant not depending on Λ. It follows that

$$
||X||_2^2\left|\left|\left(X^T\Lambda X + aX^TX\right)^{-1}\right|\right|_2^2||\mathbf{c}||_2^2 \leq K_2^2\,||X||_2^2\,||\mathbf{c}||_2^2 := K_3, \qquad \text{(B.6)}
$$

where $K_3 > 0$ is a constant not depending on Λ. Combining (B.5) and (B.6), we get

$$
\begin{aligned}
&\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right)\right|\right|_2^2 \\
&\leq \left(1 + \frac{1}{d^2}\right)\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 + \left(1 + d^2\right)K_3. \qquad \text{(B.7)}
\end{aligned}
$$

Now, from (B.1) and (B.3),

$$
\int_{\mathbb{R}^p} V(\boldsymbol{\beta})k\left(\boldsymbol{\beta}_0, \boldsymbol{\beta}\right)d\boldsymbol{\beta}
$$

$$\leq \left(1+\frac{1}{d^2}\right) E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda \mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] + \left(1+d^2\right) K_3 + a^{-1}p$$

[By (B.4) and (B.7)]

$$= \left(1+\frac{1}{d^2}\right) E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda \mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] + K_4, \qquad \text{(B.8)}$$

where $K_4 = \left(1+d^2\right) K_3 + a^{-1}p$. Again,

$$\begin{aligned}
&E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda \mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= E\left[\mathbf{z}^T\Lambda X\left(X^T\Lambda X + aX^TX\right)^{-1} \right. \\
&\qquad\qquad \left. X^TX\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= \frac{1}{a}\, E\left[\mathbf{z}^T\Lambda X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(aX^TX\right)\right. \\
&\qquad\qquad \left.\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right]. \qquad \text{(B.9)}
\end{aligned}$$

Since

$$\begin{aligned}
&\left(X^T\Lambda X + aX^TX\right)^{-1}\left(aX^TX\right)\left(X^T\Lambda X + aX^TX\right)^{-1} \\
&\preceq \left(X^T\Lambda X + aX^TX\right)^{-1}\left(X^T\Lambda X + aX^TX\right)\left(X^T\Lambda X + aX^TX\right)^{-1} \\
&= \left(X^T\Lambda X + aX^TX\right)^{-1},
\end{aligned}$$

it follows from (B.9) that

$$\begin{aligned}
&E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda \mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= \frac{1}{a}\, E\left[\mathbf{z}^T\Lambda X\left(X^T\Lambda X + aX^TX\right)^{-1}\left(aX^TX\right)\right. \\
&\qquad\qquad \left.\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&\leq \frac{1}{a}\, E\left[\mathbf{z}^T\Lambda X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right].
\end{aligned}$$

Note that $\Lambda^{\frac{1}{2}} X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda^{\frac{1}{2}} \preceq I$. It follows that

$$\begin{aligned}
&E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda \mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&\leq \frac{1}{a}\, E\left[\mathbf{z}^T\Lambda X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&\leq \frac{1}{a}\, E\left[\mathbf{z}^T\Lambda X\left(X^T\Lambda X\right)^{-1} X^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
&\leq \frac{1}{a}\, E\left[\mathbf{z}^T\Lambda\mathbf{z} \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{a} \sum_{i=1}^{n} E\left[\lambda_i z_i^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= \frac{1}{a} \sum_{i=1}^{n} E\left[E\left[\lambda_i z_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 E\left[\lambda_i \,\middle|\, z_i, \boldsymbol{\beta}_0\right] \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 \, \frac{(\nu+1)/2}{\left(\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2\right)/2} \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right]. \qquad \text{(B.10)}
\end{aligned}
$$

However, $\frac{(\nu+1)/2}{\left(\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2\right)/2} = \frac{\nu+1}{\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2} \leq \frac{\nu+1}{\nu}$. Hence

$$
\begin{aligned}
& E\left[\left|\left| X \left(X^T \Lambda X + a X^T X\right)^{-1} X^T \Lambda \mathbf{z} \right|\right|_2^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&\leq \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 \, \frac{(\nu+1)/2}{\left(\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2\right)/2} \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&\leq \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 \, \frac{\nu+1}{\nu} \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] \\
&= \frac{\nu+1}{\nu a} \sum_{i=1}^{n} E\left[z_i^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right]. \qquad \text{(B.11)}
\end{aligned}
$$

Now, from Appendix B in [16], we know that for all $\boldsymbol{\beta}_0 \in \mathbb{R}^p$,

$$
\begin{aligned}
\sum_{i=1}^{n} E\left[z_i^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right] &\leq \frac{n\nu}{\nu-2} + n\kappa M + \rho' V\left(\boldsymbol{\beta}_0\right) \\
&= \rho' V\left(\boldsymbol{\beta}_0\right) + K_5, \qquad \text{(B.12)}
\end{aligned}
$$

where κ and M are as defined in Appendix A and B respectively in [16], and $K_5 = \frac{n\nu}{\nu-2} + n\kappa M$, is a fixed constant. Combining (B.8), (B.11), and (B.12), we get

$$
\begin{aligned}
& \int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k\left(\boldsymbol{\beta}_0, \boldsymbol{\beta}\right) d\boldsymbol{\beta} \\
&\leq \left(1 + \frac{1}{d^2}\right)\left(\frac{\nu+1}{\nu a}\right)\left(\rho' V\left(\boldsymbol{\beta}_0\right) + K_5\right) + K_4 \\
&= \rho' \left(1 + \frac{1}{d^2}\right)\left(\frac{\nu+1}{\nu a}\right) V\left(\boldsymbol{\beta}_0\right) + L_1 \\
&= \rho_1 \left(1 + \frac{1}{d^2}\right) V\left(\boldsymbol{\beta}_0\right) + L_1, \qquad \text{(B.13)}
\end{aligned}
$$

where $\rho_1 = \rho'\left(\frac{\nu+1}{\nu a}\right) \geq 0$ and $L_1 > 0$ is a fixed constant. The above inequality establishes that V is a drift function if $\rho_1 < 1$ (since d is arbitrary). However, a different method of bounding various terms leads to an alternative bound that we now derive. Note that

$$\begin{aligned}
& E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
= \; & E\left[\left|\left|X\left(X^T(\Lambda + aI)X\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
= \; & E\Bigg[\bigg|\bigg|(\Lambda + aI)^{-\frac{1}{2}}(\Lambda + aI)^{\frac{1}{2}}X\left(X^T(\Lambda + aI)X\right)^{-1}X^T \\
& \qquad (\Lambda + aI)^{\frac{1}{2}}(\Lambda + aI)^{-\frac{1}{2}}\Lambda\mathbf{z}\bigg|\bigg|_2^2 \,\bigg|\boldsymbol{\beta}_0, \mathbf{y}\Bigg] \\
= \; & E\left[\left|\left|(\Lambda + aI)^{-\frac{1}{2}}P_B(\Lambda + aI)^{-\frac{1}{2}}\Lambda\mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
& \Big[\text{where, } B = (\Lambda + aI)^{\frac{1}{2}}X, \\
& \quad \text{and } P_B = B(B^TB)^{-1}B^T \text{ is the projection matrix onto } \mathcal{C}(B).\Big] \\
\leq \; & E\left[\left|\left|(\Lambda + aI)^{-\frac{1}{2}}\right|\right|_2^2 \left|\left|P_B(\Lambda + aI)^{-\frac{1}{2}}\Lambda\mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
\leq \; & \frac{1}{a}\,E\left[\left|\left|P_B(\Lambda + aI)^{-\frac{1}{2}}\Lambda\mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
\leq \; & \frac{1}{a}\,E\left[\left|\left|(\Lambda + aI)^{-\frac{1}{2}}\Lambda\mathbf{z}\right|\right|_2^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
& [\text{since for any column vector } \mathbf{w}, \\
& \quad ||P_B\mathbf{w}||_2^2 = \mathbf{w}^TP_B^TP_B\mathbf{w} = \mathbf{w}^TP_B\mathbf{w} \leq \mathbf{w}^T\mathbf{w} = ||\mathbf{w}||_2^2.] \\
= \; & \frac{1}{a}\,E\left[\sum_{i=1}^n \frac{\lambda_i^2}{\lambda_i + a}z_i^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
= \; & \frac{1}{a}\,\sum_{i=1}^n E\left[\frac{\lambda_i^2}{\lambda_i + a}z_i^2 \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
= \; & \frac{1}{a}\,\sum_{i=1}^n E\left[E\left[\frac{\lambda_i^2}{\lambda_i + a}z_i^2 \,\middle|z_i, \boldsymbol{\beta}_0\right] \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
= \; & \frac{1}{a}\,\sum_{i=1}^n E\left[z_i^2E\left[\frac{\lambda_i^2}{\lambda_i + a} \,\middle|z_i, \boldsymbol{\beta}_0\right] \,\middle|\boldsymbol{\beta}_0, \mathbf{y}\right].
\end{aligned} \tag{B.14}$$

Now,

$$\begin{aligned}
& E\left[\frac{\lambda_i^2}{\lambda_i + a} \,\middle|z_i, \boldsymbol{\beta}_0\right] \\
= \; & E\left[\lambda_i\frac{\lambda_i}{\lambda_i + a} \,\middle|z_i, \boldsymbol{\beta}_0\right]
\end{aligned}$$

$$\leq \sqrt{E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] E\left[\left(\frac{\lambda_i}{\lambda_i + a}\right)^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right]}$$

[by Cauchy-Schwarz inequality]

$$\leq \sqrt{E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] E\left[\frac{\lambda_i^2}{\lambda_i^2 + a^2} \,\middle|\, z_i, \boldsymbol{\beta}_0\right]}. \tag{B.15}$$

Since the function $f(x) = \frac{x}{x+a^2}$ is a concave function in x ($\frac{d^2}{dx^2} f(x) < 0$), applying Jensen's inequality gives

$$E\left[\frac{\lambda_i^2}{\lambda_i^2 + a^2} \,\middle|\, z_i, \boldsymbol{\beta}_0\right] \leq \frac{E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right]}{E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] + a^2}. \tag{B.16}$$

Combining (B.15) and (B.16), we get

$$E\left[\frac{\lambda_i^2}{\lambda_i + a} \,\middle|\, z_i, \boldsymbol{\beta}_0\right] \leq \sqrt{E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] \frac{E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right]}{\left(E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] + a^2\right)}}. \tag{B.17}$$

Since $\lambda_i \Big| z_i, \boldsymbol{\beta}_0 \sim \text{Gamma}\left(\frac{\nu+1}{2}, \frac{\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2}{2}\right)$, we have

$$E\left[\lambda_i^2 \,\middle|\, z_i, \boldsymbol{\beta}_0\right] = \frac{(\nu+1)(\nu+3)}{\left(\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2\right)^2} \leq \frac{(\nu+1)(\nu+3)}{\nu^2}. \tag{B.18}$$

Combining (B.17) and (B.18),

$$E\left[\frac{\lambda_i^2}{\lambda_i + a} \,\middle|\, z_i, \boldsymbol{\beta}_0\right]$$

$$\leq \sqrt{\frac{(\nu+1)(\nu+3)}{\nu^2} \frac{\frac{(\nu+1)(\nu+3)}{\nu^2}}{\left(\frac{(\nu+1)(\nu+3)}{\nu^2} + a^2\right)}}$$

$$\left[\text{since } x \mapsto \frac{x}{x+a^2} \text{ is an increasing function in } x\right]$$

$$= \frac{(\nu+1)(\nu+3)}{\nu^2} \sqrt{\frac{\nu^2}{(\nu+1)(\nu+3) + a^2\nu^2}}$$

$$= \frac{(\nu+1)}{\nu} \sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) + a^2\nu^2}}. \tag{B.19}$$

Now, combining (B.14) and (B.19), we get

$$E\left[\left|\left|X\left(X^T \Lambda X + a X^T X\right)^{-1} X^T \Lambda \mathbf{z}\right|\right|_2^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y}\right]$$

$$
\begin{aligned}
&\leq \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 E\left[\frac{\lambda_i^2}{\lambda_i + a} \,\middle|\, z_i, \boldsymbol{\beta}_0 \right] \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y} \right] \\
&\leq \frac{(\nu+1)}{\nu a} \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right) \sum_{i=1}^{n} E\left[z_i^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y} \right] \\
&\leq \frac{(\nu+1)}{\nu a} \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right) \left(\rho' V(\boldsymbol{\beta}_0) + K_5 \right) \qquad \text{[by (B.12)]} \\
&= \rho' \left(\frac{\nu+1}{\nu a} \right) \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right) V(\boldsymbol{\beta}_0) \;+\; \tilde{K}, \qquad \text{(B.20)}
\end{aligned}
$$

where $\tilde{K} = K_5 \left(\frac{\nu+1}{\nu a}\right) \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right)$ is a fixed constant. Finally, combining (B.8) and (B.20), we get

$$
\begin{aligned}
&\int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k(\boldsymbol{\beta}_0, \boldsymbol{\beta})\, d\boldsymbol{\beta} \\
&= \left(1 + \frac{1}{d^2}\right) E\left[\left\| X \left(X^T \Lambda X + a X^T X \right)^{-1} X^T \Lambda \mathbf{z} \right\|_2^2 \,\middle|\, \boldsymbol{\beta}_0, \mathbf{y} \right] \;+\; K_4 \\
&\leq \left(1 + \frac{1}{d^2}\right) \left[\rho' \left(\frac{\nu+1}{\nu a} \right) \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right) V(\boldsymbol{\beta}_0) \right. \\
&\qquad\qquad \left. + \; \tilde{K} \right] \;+\; K_4 \\
&= \left[\rho' \left(1 + \frac{1}{d^2}\right) \left(\frac{\nu+1}{\nu a} \right) \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right) \right] V(\boldsymbol{\beta}_0) \;+\; L_2 \\
&= \rho_2 \left(1 + \frac{1}{d^2}\right) V(\boldsymbol{\beta}_0) \;+\; L_2, \qquad \text{(B.21)}
\end{aligned}
$$

where $\rho_2 = \rho' \left(\frac{\nu+1}{\nu a}\right) \left(\sqrt{\frac{(\nu+3)^2}{(\nu+1)(\nu+3) \;+\; a^2\nu^2}} \right) \geq 0$ and $L_2 > 0$ is a fixed constant. From (B.13) and (B.21), we conclude

$$
\int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k(\boldsymbol{\beta}_0, \boldsymbol{\beta})\, d\boldsymbol{\beta} \;\leq\; \min\{\rho_1, \rho_2\} \left(1 + \frac{1}{d^2}\right) V(\boldsymbol{\beta}_0) \;+\; \max\{L_1, L_2\}.
$$

Since $\min\{\rho_1, \rho_2\} < 1$ by assumption, $d > 0$ is an arbitrary real number which can be chosen large enough so that $\left[\min\{\rho_1, \rho_2\} \left(1 + \frac{1}{d^2}\right)\right]$ is less than 1. This establishes a geometric drift condition for the function V. Since V is unbounded off compact sets, a standard argument using Fatou's lemma along with Theorem 6.0.1 and Lemma 15.2.8 of [11] implies geometric ergodicity of the robit DA chain. □

Appendix C: Geometric ergodicity of the robit DA chain using drift and minorization in the high-dimensional setting (with $n < p$)

We now establish geometric ergodicity of the robit DA chain in a high-dimensional setting with $n < p$. This setting was not considered in [16]. Let

$$\rho_2' = \left(\frac{\frac{(\nu+1)(\nu+3)}{\nu^2}}{\frac{(\nu+1)(\nu+3)}{\nu^2} + a^2} \right).$$

Theorem C.1. *Suppose $n < p$. Then, the robit DA Markov chain is geometrically ergodic if*

(A) $\nu > 2$,
(B) $\Sigma_a = aX^TX + \Sigma$ *is positive definite, where* Σ *is non-negative definite with* $X\Sigma = 0$.

We briefly comment on the similarities and differences between the assumptions for Theorem B.1 and those for Theorem C.1. Both results require $\nu > 2$ and both the corresponding proofs involve establishing a geometric drift condition for $V(\boldsymbol{\beta}) = \boldsymbol{\beta}^T X^T X \boldsymbol{\beta}$. The required form of the prior precision matrix Σ_a is also slightly different in the two results, partly because X^TX is not non-singular in the $n < p$ setting. Finally, we will establish a geometric drift condition for V in the proof of Theorem C.1 where the coefficient of V on the R.H.S. is $\min(\rho_1, \rho_2')$. Here

$$\rho_2' = \rho' \left(\frac{\frac{(\nu+1)(\nu+3)}{\nu^2}}{\frac{(\nu+1)(\nu+3)}{\nu^2} + a^2} \right) < 1.$$

Hence an analogue of assumption (C) in Theorem B.1 is not required for Theorem C.1.

Proof. We will use the same function $V : \mathbb{R}^p \to \mathbb{R}_+$ as in Appendix B, i.e., $V(\boldsymbol{\beta}) = \boldsymbol{\beta}^T X^T X \boldsymbol{\beta}$. Since $n < p$, the design matrix X no longer has full column rank and therefore, the level set $\{\boldsymbol{\beta} : V(\boldsymbol{\beta}) \leq \alpha\}$ may not be compact for some $\alpha > 0$. In other words, V is not unbounded off compact sets in this setting, and we will have to establish an additional minorization condition on top of a geometric drift condition for V to establish geometric ergodicity.

We first establish a geometric drift condition for V. Recall that the robit DA Markov chain transitions from the state $\boldsymbol{\beta}_0$ to the state $\boldsymbol{\beta}$ through the following two steps i.e. by generating the intermediate latent variables $(\boldsymbol{\lambda}, \mathbf{z})$ first and then generating $\boldsymbol{\beta}$ from the relevant conditional distribution. Now, following the same arguments as in (B.1) in Appendix B, we note that

$$\int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k(\boldsymbol{\beta}_0, \boldsymbol{\beta})\, d\boldsymbol{\beta} \;=\; E\left[E\left[||X\boldsymbol{\beta}||_2^2 \;\middle|\; \boldsymbol{\lambda}, \mathbf{z}, \mathbf{y} \right] \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y} \right], \tag{C.1}$$

where the inner expectation is w.r.t. the conditional distribution of $\boldsymbol{\beta}$ given $(\boldsymbol{\lambda}, \mathbf{z}, \mathbf{y})$, and the outer expectation is w.r.t. the conditional distribution of $(\boldsymbol{\lambda}, \mathbf{z})$ given $(\boldsymbol{\beta}_0, \mathbf{y})$.

Using (B.2), the inner expectation in (C.1) is given by

$$E\left(||X\boldsymbol{\beta}||_2^2 \,\middle|\, \boldsymbol{\lambda}, \mathbf{z}, \mathbf{y}\right) = \left|\left|X\left(X^T\Lambda X + \Sigma_a\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right)\right|\right|_2^2 + \text{tr}\left(X\left(X^T\Lambda X + \Sigma_a\right)^{-1}X^T\right). \tag{C.2}$$

Since $n < p$, the singular value decomposition of X is given by

$$X_{n\times p} = V_{n\times n}D_{n\times n}U_{n\times p}^T,$$

where $U_{p\times n}$ is a semi-orthogonal matrix with full column rank n, V is an orthogonal matrix and D is a diagonal matrix with singular values of X being its diagonal entries. Let $\tilde{U}_{p\times(p-n)}$ be a semi-orthogonal matrix whose column space is the orthogonal complement of the column space of U. Hence, the matrix $\begin{bmatrix}U & \tilde{U}\end{bmatrix}_{p\times p}$ is an orthogonal matrix. Since $X\Sigma = 0$ by Assumption (B), it follows that $\Sigma = \tilde{U}\tilde{D}\tilde{U}^T$ for some diagonal matrix $\tilde{D}$, and

$$\Sigma_a \;=\; \begin{bmatrix}U & \tilde{U}\end{bmatrix}\begin{bmatrix}aD^2 & 0\\ 0 & \tilde{D}\end{bmatrix}\begin{bmatrix}U^T\\ \tilde{U}^T\end{bmatrix}.$$

Note that $a > 0$ and $D, \tilde{D}$ have positive diagonal entries as Σ_a is assumed to be positive definite. Using the fact that V and $\begin{bmatrix}U & \tilde{U}\end{bmatrix}_{p\times p}$ are orthogonal matrices, the second-term in the R.H.S of (C.2) can be bounded as

$$\begin{aligned}
&\text{tr}\left(X\left(X^T\Lambda X + \Sigma_a\right)^{-1}X^T\right)\\
&= \;\text{tr}\left(DU^T\begin{bmatrix}U^T\\ \tilde{U}^T\end{bmatrix}^{-1}\begin{bmatrix}DV^T\Lambda VD + aD^2 & 0\\ 0 & \tilde{D}\end{bmatrix}^{-1}\begin{bmatrix}U & \tilde{U}\end{bmatrix}^{-1}UD\right)\\
&= \;\text{tr}\left(\begin{bmatrix}D & 0\end{bmatrix}\begin{bmatrix}\left(DV^T\Lambda VD + aD^2\right)^{-1} & 0\\ 0 & \tilde{D}^{-1}\end{bmatrix}\begin{bmatrix}D\\ 0\end{bmatrix}\right)\\
&= \;\text{tr}\left(D\left(DV^T\Lambda VD + aD^2\right)^{-1}D\right)\\
&= \;\text{tr}\left(\left(V^T\Lambda V + aI_n\right)^{-1}\right) \le \text{tr}\left(a^{-1}I_n\right) = a^{-1}n.
\end{aligned}$$

For the first-term in the R.H.S. of (C.2), we follow the same arguments as in (B.5) in Appendix B to get

$$\begin{aligned}
&\left|\left|X\left(X^T\Lambda X + \Sigma_a\right)^{-1}\left(X^T\Lambda\mathbf{z} + \mathbf{c}\right)\right|\right|_2^2\\
&\le \left(1 + \frac{1}{d^2}\right)\left|\left|X\left(X^T\Lambda X + \Sigma_a\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2\\
&\qquad\qquad + \left(1 + d^2\right)||X||_2^2\left|\left|\left(X^T\Lambda X + \Sigma_a\right)^{-1}\right|\right|_2^2||\mathbf{c}||_2^2\\
&\le \left(1 + \frac{1}{d^2}\right)\left|\left|X\left(X^T\Lambda X + \Sigma_a\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2
\end{aligned}$$

$$+\left(1+d^2\right)||X||_2^2\left|\left|\Sigma_a^{-1}\right|\right|_2^2||\mathbf{c}||_2^2, \tag{C.3}$$

where $d>0$ is an arbitrary real number. Now, from (C.1), (C.2) and (C.3), it follows that

$$\begin{aligned}&\int_{\mathbb{R}^p} V(\boldsymbol{\beta})k\left(\boldsymbol{\beta}_0,\boldsymbol{\beta}\right)d\boldsymbol{\beta}\\ &=\left(1+\frac{1}{d^2}\right)E\left[\left|\left|X\left(X^T\Lambda X+\Sigma_a\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right]+K_4,\end{aligned} \tag{C.4}$$

where K_4 is a fixed constant. Again, using the fact that V and $\begin{bmatrix}U & \tilde{U}\end{bmatrix}_{p\times p}$ are orthogonal matrices, we get

$$\begin{aligned}&\left|\left|X\left(X^T\Lambda X+\Sigma_a\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2\\ =&\ \left|\left|VDU^T\left(UDV^T\Lambda VDU^T+U\left(aD^2\right)U^T+\tilde{U}\tilde{D}\tilde{U}^T\right)^{-1}UDV^T\Lambda\mathbf{z}\right|\right|_2^2\\ =&\ \left|\left|VD\begin{bmatrix}I_n & 0\end{bmatrix}\begin{bmatrix}\left(DV^T\Lambda VD+aD^2\right)^{-1} & 0\\ 0 & \tilde{D}^{-1}\end{bmatrix}\begin{bmatrix}I_n\\ 0\end{bmatrix}DV^T\Lambda\mathbf{z}\right|\right|_2^2\\ =&\ \left|\left|V\left(V^T\Lambda V+aI_n\right)^{-1}V^T\Lambda\mathbf{z}\right|\right|_2^2\\ =&\ \left|\left|\left(\Lambda+aI_n\right)^{-1}\Lambda\mathbf{z}\right|\right|_2^2\\ =&\ \sum_{i=1}^n\left(\frac{\lambda_i}{\lambda_i+a}\right)^2z_i^2.\end{aligned} \tag{C.5}$$

Hence,

$$\begin{aligned}&E\left[\left|\left|X\left(X^T\Lambda X+\Sigma_a\right)^{-1}X^T\Lambda\mathbf{z}\right|\right|_2^2\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right]\\ =&\ E\left[\sum_{i=1}^n\left(\frac{\lambda_i}{\lambda_i+a}\right)^2z_i^2\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right] \qquad \text{[from (C.5)]}\\ =&\ \sum_{i=1}^n E\left[\left(\frac{\lambda_i}{\lambda_i+a}\right)^2z_i^2\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right]\\ \le&\ \sum_{i=1}^n E\left[E\left[\frac{\lambda_i^2}{\lambda_i^2+a^2}\,z_i^2\ \Big|\ z_i,\boldsymbol{\beta}_0\right]\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right]\\ \le&\ \sum_{i=1}^n E\left[z_i^2\,\frac{E\left[\lambda_i^2\ \big|z_i,\boldsymbol{\beta}_0\right]}{E\left[\lambda_i^2\ \big|z_i,\boldsymbol{\beta}_0\right]+a^2}\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right] \qquad \text{[By (B.16) in Appendix } B\text{]}\\ \le&\ \frac{\frac{(\nu+1)(\nu+3)}{\nu^2}}{\left(\frac{(\nu+1)(\nu+3)}{\nu^2}+a^2\right)}\sum_{i=1}^n E\left[z_i^2\ \Big|\ \boldsymbol{\beta}_0,\mathbf{y}\right].\end{aligned} \tag{C.6}$$

$$[\text{By (B.18) in Appendix } B$$

$$\left.\text{and since } x \mapsto \frac{x}{x+a^2} \text{ is an increasing function in } x\right]$$

From Appendix B in [16], we know that

$$\begin{aligned}
\sum_{i=1}^{n} E\left[z_i^2 \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right] &= \frac{n\nu}{\nu-2} + \sum_{i=1}^{n}\left(\mathbf{w}_i^T \boldsymbol{\beta}_0\right)^2 \\
&\quad - \kappa \sum_{i=1}^{n} \frac{\mathbf{w}_i^T \boldsymbol{\beta}_0}{1 - F_\nu\left(\mathbf{w}_i^T \boldsymbol{\beta}_0\right)} \cdot \frac{1}{\left(\nu + \left(\mathbf{w}_i^T \boldsymbol{\beta}_0\right)^2\right)^{\frac{\nu-1}{2}}} \\
&\leq \rho' V\left(\boldsymbol{\beta}_0\right) + \frac{n\nu}{\nu-2} + n\kappa M, \qquad \text{(C.7)}
\end{aligned}$$

where ρ', κ, M are defined in Appendix B of [16]. It follows from (C.4), (C.6) and (C.7) that for all $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, we have

$$\int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k\left(\boldsymbol{\beta}_0, \boldsymbol{\beta}\right) d\boldsymbol{\beta} = \left(1 + \frac{1}{d^2}\right) \rho_2' V\left(\boldsymbol{\beta}_0\right) \; + \; L, \qquad \text{(C.8)}$$

where $L = \left(1+\frac{1}{d^2}\right)\left(\frac{\frac{(\nu+1)(\nu+3)}{\nu^2}}{\frac{(\nu+1)(\nu+3)}{\nu^2}+a^2}\right)\left(\frac{n\nu}{\nu-2} + n\kappa M\right) + K_4 \;>\; 0$.Since $\rho_2' < 1$ and $d > 0$ is arbitrary, the above already provides a geometric drift condition. However, we can tighten our analysis by using an alternative bound instead of (C.6). Note that

$$\left(\frac{\lambda_i}{\lambda_i + a}\right)^2 \leq \frac{\lambda_i}{\lambda_i + a} \leq \frac{\lambda_i}{a}.$$

It follows from (C.5) that

$$\begin{aligned}
& E\left[\left|\left|X\left(X^T \Lambda X + \Sigma_a\right)^{-1} X^T \Lambda \mathbf{z}\right|\right|_2^2 \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right] \\
\leq\; & E\left[a^{-1} \mathbf{z}^T \Lambda \mathbf{z} \;\middle|\; \boldsymbol{\beta}_0, \mathbf{y}\right] \\
=\; & \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 \, \frac{(\nu+1)/2}{\left(\nu + \left(z_i - \mathbf{x}_i^T \boldsymbol{\beta}_0\right)^2\right)/2} \;\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
& \qquad [\text{By (B.10) in Appendix B}] \\
\leq\; & \frac{1}{a} \sum_{i=1}^{n} E\left[z_i^2 \, \frac{\nu+1}{\nu} \;\middle|\boldsymbol{\beta}_0, \mathbf{y}\right] \\
=\; & \frac{\nu+1}{\nu a} \sum_{i=1}^{n} E\left[z_i^2 \;\middle|\boldsymbol{\beta}_0, \mathbf{y}\right]. \qquad \text{(C.9)}
\end{aligned}$$

Using the definition of ρ' in Appendix B of [16], and combining (C.4) and (C.9), we get

$$\int_{\mathbb{R}^p} V(\boldsymbol{\beta}) k\left(\boldsymbol{\beta}_0, \boldsymbol{\beta}\right) d\boldsymbol{\beta}$$

$$
\begin{aligned}
&\le \left(1+\frac{1}{d^2}\right)\left(\frac{\nu+1}{\nu a}\right)\left(\rho' V(\boldsymbol\beta_0)+\frac{n\nu}{\nu-2}+n\kappa M\right) + K_4 \\
&= \rho'\left(1+\frac{1}{d^2}\right)\left(\frac{\nu+1}{\nu a}\right)V(\boldsymbol\beta_0) + L_1 \\
&= \rho_1\left(1+\frac{1}{d^2}\right)V(\boldsymbol\beta_0) + L_1, \qquad \text{(C.10)}
\end{aligned}
$$

where $\rho_1=\rho'\left(\frac{\nu+1}{\nu a}\right)$ and $L_1=\left(1+\frac{1}{d^2}\right)\left(\frac{\nu+1}{\nu a}\right)\left(\frac{n\nu}{\nu-2}+n\kappa M\right)+K_4 > 0$.
It follows from (C.8) and (C.10) that

$$
\int_{\mathbb{R}^p} V(\boldsymbol\beta)k(\boldsymbol\beta_0,\boldsymbol\beta)\,d\boldsymbol\beta \le \min\{\rho_1,\rho_2'\}\left(1+\frac{1}{d^2}\right)V(\boldsymbol\beta_0) + \max\{L_1,L_2\}.
$$

Since $\min\{\rho_1,\rho_2'\} < 1$ by assumption, and $d > 0$ is arbitrary, the required geometric drift condition follows.

Since V is not unbounded off compact sets in this setting, we will now establish an associated minorization condition, i.e., for any $d>0$, establish the existence of an $\epsilon=\epsilon(d)>0$ and a probability density h such that $k(\boldsymbol\beta_0,\boldsymbol\beta)\ge\epsilon h(\boldsymbol\beta)$ whenever $V(\boldsymbol\beta_0)\le d$. Recall that the transition density k of the robit DA chain is given by

$$
\begin{aligned}
k(\boldsymbol\beta_0,\boldsymbol\beta) &= \int_{\mathbb{R}_+^n}\int_{\mathbb{R}^n}\pi\left(\boldsymbol\beta\middle|\boldsymbol\lambda,\mathbf{z},\mathbf{y}\right)\pi\left(\boldsymbol\lambda,\mathbf{z}\middle|\boldsymbol\beta_0,\mathbf{y}\right)d\mathbf{z}\,d\boldsymbol\lambda \\
&= \int_{\mathbb{R}_+^n}\int_{\mathbb{R}^n}\pi\left(\boldsymbol\beta\middle|\boldsymbol\lambda,\mathbf{z},\mathbf{y}\right)\pi\left(\boldsymbol\lambda\middle|\mathbf{z},\boldsymbol\beta_0,\mathbf{y}\right)\pi\left(\mathbf{z}\middle|\boldsymbol\beta_0,\mathbf{y}\right)d\mathbf{z}\,d\boldsymbol\lambda.
\end{aligned}
$$

Note that to show the required minorization condition, it is enough to show that whenever $V(\boldsymbol\beta_0)\le d$ we have

$$
\pi\left(\boldsymbol\lambda,\mathbf{z}\middle|\boldsymbol\beta_0,\mathbf{y}\right) \ge \epsilon\,\tilde h(\boldsymbol\lambda,\mathbf{z}), \qquad \text{(C.11)}
$$

for some real number $\epsilon > 0$ and a probability density $\tilde h$. The minorization condition will then hold with the choice

$$
h(\boldsymbol\beta)=\int_{\mathbb{R}_+^n}\int_{\mathbb{R}^n}\pi\left(\boldsymbol\beta\middle|\boldsymbol\lambda,\mathbf{z},\mathbf{y}\right)\tilde h(\boldsymbol\lambda,\mathbf{z})d\mathbf{z}\,d\boldsymbol\lambda.
$$

Note that if $V(\boldsymbol\beta_0)\le d$, then $\left(\mathbf{x}_i^T\boldsymbol\beta_0\right)^2\le d$ for every $1\le i\le n$. Hence, for any such $\boldsymbol\beta_0$, we get

$$
\begin{aligned}
\pi\left(\boldsymbol\lambda\middle|\mathbf{z},\boldsymbol\beta_0,\mathbf{y}\right) &= \prod_{i=1}^n \frac{\left(\frac{\nu+\left(z_i-\mathbf{x}_i^T\boldsymbol\beta_0\right)^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)}\lambda_i^{\frac{\nu-1}{2}}\, e^{-\frac{\lambda_i}{2}\left(\nu+\left(z_i-\mathbf{x}_i^T\boldsymbol\beta_0\right)^2\right)} \\
&\ge \prod_{i=1}^n \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)}\lambda_i^{\frac{\nu-1}{2}}\, e^{-\frac{\lambda_i}{2}\left(\nu+2z_i^2+2\left(\mathbf{x}_i^T\boldsymbol\beta_0\right)^2\right)}
\end{aligned}
$$

$$
\begin{aligned}
&\geq \prod_{i=1}^{n} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \lambda_i^{\frac{\nu-1}{2}} \; e^{-\frac{\lambda_i}{2}\left(\nu+2z_i^2+2d\right)} \\
&= \prod_{i=1}^{n} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \lambda_i^{\frac{\nu-1}{2}} \; e^{-\lambda_i\left(\frac{\nu}{2}+z_i^2+d\right)} \\
&= \left(\frac{\nu}{2}\right)^{n\left(\frac{\nu+1}{2}\right)} \left(\prod_{i=1}^{n} \left(\frac{\nu}{2}+z_i^2+d\right)^{-\frac{\nu+1}{2}}\right) g\left(\boldsymbol{\lambda}\middle|\mathbf{z},\mathbf{y}\right),
\end{aligned}
\tag{C.12}
$$

where

$$
g\left(\boldsymbol{\lambda}\middle|\mathbf{z},\mathbf{y}\right) \;=\; \prod_{i=1}^{n} \frac{\left(\frac{\nu}{2}+z_i^2+d\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \lambda_i^{\frac{\nu-1}{2}} \; e^{-\lambda_i\left(\frac{\nu}{2}+z_i^2+d\right)},
$$

is the product of n i.i.d Gamma$\left(\frac{(\nu+1)}{2}, \left(\frac{\nu}{2}+z_i^2+d\right)\right)$ densities.

Moreover, if we denote $t_\nu\left(\mu,1\right)$ to be a random variable having univariate Student's t distribution with location μ, scale 1 and degrees of freedom ν, then for for any $i \in \{1,2,\ldots,n\}$ with $y_i = 0$, we get

$$
\begin{aligned}
\pi\left(z_i\middle|\boldsymbol{\beta}_0, y_i\right) \;&=\; \frac{c\left(1+\frac{\left(z_i-\mathbf{x}_i^T\boldsymbol{\beta}_0\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(\mathbf{x}_i^T\boldsymbol{\beta}_0,1\right)\leq 0\right)} \\
&\qquad \text{[where } c \text{ is the normalizing constant.]} \\
&=\; \frac{c\left(1+\frac{\left(z_i-\mathbf{x}_i^T\boldsymbol{\beta}_0\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\leq -\mathbf{x}_i^T\boldsymbol{\beta}_0\right)} \\
&\geq\; \frac{c\left(1+\frac{2z_i^2+2\left(\mathbf{x}_i^T\boldsymbol{\beta}_0\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\leq -\mathbf{x}_i^T\boldsymbol{\beta}_0\right)} \\
&\geq\; \frac{c\left(1+\frac{2z_i^2+2d}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\leq \sqrt{d}\right)} \\
&=\; \frac{c\left(\frac{\nu}{2}\right)^{\frac{\nu+1}{2}}\left(\frac{\nu}{2}+z_i^2+d\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\leq \sqrt{d}\right)},
\end{aligned}
\tag{C.13}
$$

whenever $V(\boldsymbol{\beta}_0) \leq d$. Similarly for any $i \in \{1,2,\ldots,n\}$ with $y_i = 1$, we get

$$\begin{aligned}
\pi\left(z_i\middle|\boldsymbol{\beta}_0, y_i\right) &= \frac{c\left(1+\frac{\left(z_i-\mathbf{x}_i^T\boldsymbol{\beta}_0\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(\mathbf{x}_i^T\boldsymbol{\beta}_0,1\right)\geq 0\right)} \\
&\qquad\text{[where } c \text{ is the same as in (C.13)]} \\
&= \frac{c\left(1+\frac{\left(z_i-\mathbf{x}_i^T\boldsymbol{\beta}_0\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\geq -\mathbf{x}_i^T\boldsymbol{\beta}_0\right)} \\
&\geq \frac{c\left(1+\frac{2z_i^2+2\left(\mathbf{x}_i^T\boldsymbol{\beta}_0\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\geq -\mathbf{x}_i^T\boldsymbol{\beta}_0\right)} \\
&\geq \frac{c\left(1+\frac{2z_i^2+2d}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\geq -\sqrt{d}\right)} \\
&= \frac{c\left(1+\frac{2z_i^2+2d}{\nu}\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\leq \sqrt{d}\right)} \\
&\qquad\text{[As } t_\nu\left(0,1\right) \text{ is symmetric around 0]} \\
&= \frac{c\left(\frac{\nu}{2}\right)^{\frac{\nu+1}{2}}\left(\frac{\nu}{2}+z_i^2+d\right)^{-\frac{\nu+1}{2}}}{P\left(t_\nu\left(0,1\right)\leq \sqrt{d}\right)}, \qquad \text{(C.14)}
\end{aligned}$$

whenever $V(\boldsymbol{\beta}_0)\leq d$. Combining (C.13) and (C.14), we obtain

$$\pi\left(\mathbf{z}\middle|\boldsymbol{\beta}_0,\mathbf{y}\right) \geq \frac{c^n\left(\frac{\nu}{2}\right)^{n\left(\frac{\nu+1}{2}\right)}}{\left[P\left(t_\nu\left(0,1\right)\leq\sqrt{d}\right)\right]^n}\prod_{i=1}^{n}\left(\frac{\nu}{2}+z_i^2+d\right)^{-\frac{\nu+1}{2}}, \qquad \text{(C.15)}$$

for every $1\leq i\leq n$ whenever $V(\boldsymbol{\beta}_0)\leq d$. Now, using (C.12) and (C.15), we have

$$\begin{aligned}
\pi\left(\boldsymbol{\lambda}\middle|\mathbf{z},\boldsymbol{\beta}_0,\mathbf{y}\right)\pi\left(\mathbf{z}\middle|\boldsymbol{\beta}_0,\mathbf{y}\right) &\geq \frac{c^n\left(\frac{\nu}{2}\right)^{n(\nu+1)}}{\left[P\left(t_\nu\left(0,1\right)\leq\sqrt{d}\right)\right]^n}\; g\left(\boldsymbol{\lambda}\middle|\mathbf{z},\mathbf{y}\right) \\
&\quad\times\prod_{i=1}^{n}\left(\frac{\nu}{2}+z_i^2+d\right)^{-(\nu+1)}. \qquad \text{(C.16)}
\end{aligned}$$

Again, for any $i\in\{1,2,\ldots,n\}$,

$$\int_{-\infty}^{\infty}\left(\frac{\nu}{2}+z_i^2+d\right)^{-(\nu+1)}dz_i$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \left(K + z_i^2\right)^{-(\nu+1)} dz_i \\
&\qquad \left[\text{where, } K = \left(\frac{\nu}{2} + d\right) \text{ is a fixed constant}\right] \\
&= K^{-(\nu+1)} \int_{-\infty}^{\infty} \left(1 + \frac{z_i^2}{K}\right)^{-(\nu+1)} dz_i \\
&= K^{-(\nu+1)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(1 + \tan^2(x)\right)^{-(\nu+1)} \frac{\sqrt{K}}{\cos^2(x)}\ dx \\
&\left[\text{by taking } \frac{z_i}{\sqrt{K}} = \tan(x) \implies dz_i = \sqrt{K}\sec^2(x)\ dx = \frac{\sqrt{K}}{\cos^2(x)}\ dx\right] \\
&= K^{-(\nu+\frac{1}{2})} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\sec^2(x)\right)^{-(\nu+1)} \frac{1}{\cos^2(x)}\ dx \\
&\qquad \left[\text{since } 1 + \tan^2(x) = \sec^2(x)\right] \\
&= K^{-(\nu+\frac{1}{2})} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{2\nu}(x)\ dx \\
&= 2\ K^{-(\nu+\frac{1}{2})} \int_{0}^{\frac{\pi}{2}} \cos^{2\nu}(x)\ dx \\
&\qquad \left[\text{since } \cos^{2\nu}(x) \text{ is an even function in } x\right] \\
&\le 2\ K^{-(\nu+\frac{1}{2})} \int_{0}^{\frac{\pi}{2}} 1\ dx \\
&\qquad \left[\text{since } \cos^2(x) \le 1 \implies \cos^{2\nu}(x) \le 1 \text{ as } \nu > 2\right] \\
&= \pi\ K^{-(\nu+\frac{1}{2})} \\
&< \infty.
\end{aligned}
$$

Hence, the lower bound in (C.16) (which does not depend on $\boldsymbol{\beta}_0$) can be normalized to a probability density to obtain ϵ and $\tilde{h}$ as required in (C.11). This establishes the minorization condition. The drift and minorization conditions above can be combined with Theorem 12 of [15] to establish geometric ergodicity of the robit DA Markov chain. □

Remark C.1. A key step, both in the proof of Theorem B.1 and of Theorem C.1 is the bounding of the expectation (see Eq. (B.10) and Eq. (C.5)):

$$E\left[\left|\left|X\left(X^T\Lambda X + aX^TX\right)^{-1} X^T\Lambda\mathbf{z}\right|\right|_2^2 \ \middle|\boldsymbol{\beta}_0, \mathbf{y}\right].$$

The expectation is with respect to the conditional distribution of $(\boldsymbol{\lambda}, \mathbf{z})$ given $\boldsymbol{\beta}_0, \mathbf{y}$. If Λ is replaced by the identity matrix (as in the probit case), then it can be shown (see [2]) that

$$\left|\left|X\left(X^TX + \Sigma_a\right)^{-1} X^T\mathbf{z}\right|\right|_2^2 \le e_{\max}\|\mathbf{z}\|_2^2$$

where $e_{\max} < 1$ if Σ_a is positive definite. This enables establishing a drift condition in the probit setting without any (additional) constraints on ν and Σ_a. However, the presence of Λ in the robit setting, and the corresponding additional expectation step, leads to modified bounds as provided in Eq. (B.11) and Eq. (C.6) in the Appendix. These modified bounds necessitate the use of appropriate assumptions on Σ_a and ν for obtaining a geometric drift condition.

Appendix D: Traceplots for MCMC chains associated with the Lupus data example

Below we show the traceplots for the Markov chain realizations for β_1 and β_2 corresponding to Bayesian robit models fitted to the Lupus dataset with $\nu = 1$, $\nu = 3$, and $\nu = 1000$.

Fig D.1. *Traceplots for the realizations of regression coefficients β_1 and β_2 obtained from the DA and sandwich algorithms for Bayesian robit model with $\nu = 1$ implemented on the Lupus dataset, using two different Zellner's g priors. The iterations on the left-hand side of the vertical bar (iteration 10,000) are treated as burn-in.*

FIG D.2. *Traceplots for the realizations of regression coefficients β_1 and β_2 obtained from the DA and sandwich algorithms for Bayesian robit model with $\nu = 3$ implemented on the Lupus dataset, using two different Zellner's g priors. The iterations on the left-hand side of the vertical bar (iteration* 10,000*) are treated as burn-in.*

FIG D.3. *Traceplots for the realizations of regression coefficients β_1 and β_2 obtained from the DA and sandwich algorithms for Bayesian robit model with $\nu = 1000$ implemented on the Lupus dataset, using two different Zellner's g priors. The iterations on the left-hand side of the vertical bar (iteration* 10,000*) are treated as burn-in.*

References

[1] James H. Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993. MR1224394

[2] Saptarshi Chakraborty and Kshitij Khare. Convergence properties of Gibbs samplers for Bayesian probit regression with proper priors. *Electronic Journal of Statistics*, 11:177–210, 2017. MR3604022

[3] Saptarshi Chakraborty and Kshitij Khare. Consistent estimation of the spectrum of trace class data augmentation algorithms. *Bernoulli*, 25:3832–3863, 2019. MR4010974

[4] Hyonho Chun and Sündüz Keleş. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25, 2010. MR2751241

[5] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, 2007.

[6] Andrew Gelman, Jennifer Hill, and Aki Vehtari. *Regression and Other Stories*. Cambridge University Press, Cambridge, 2020.

[7] K Jörgens. Linear integral operators. Surveys and reference works in mathematics. *Pitman Advanced Pub. Program*, 1982. MR0647629

[8] K. Khare and J. P. Hobert. A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *The Annals of Statistics*, 39:2585–2606, 2011. MR2906879

[9] Chuanhai Liu. *Robit Regression: A Simple Robust Alternative to Logistic and Probit Regression*, chapter 21, pages 227–238. John Wiley & Sons, Ltd, 2004. MR2138259

[10] Jun S. Liu, Wing Hung Wong, and Augustine Kong. Covariance structure of the gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81(1):27–40, 1994. MR1279653

[11] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993. MR1287609

[12] N. Balakrishnan N.L. Johnson, S. Kotz. Continuous Univariate Distributions, Volume 2, 2nd Edition. *Wiley*, 1995. MR1326603

[13] Daryl Pregibon. Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, 38(2):485–498, 1982.

[14] Qian Qin, James P. Hobert, and Kshitij Khare. Estimating the spectral gap of a trace-class Markov operator. *Electronic Journal of Statistics*, 13:1790–1822, 2019. MR3959873

[15] J.S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558–566, 1995. MR1340509

[16] Vivekananda Roy. Convergence rates for MCMC algorithms for a robust Bayesian binary regression model. *Electronic Journal of Statistics*, 6(none):2463–2485, 2012. MR3020272

[17] Vivekananda Roy and James P. Hobert. Convergence rates and asymptotic standard errors for markov chain monte carlo algorithms for bayesian probit regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(4):607–623, 2007. MR2370071

[18] David A. van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50, 2001. MR1936358