GENERATIVE COLLAPSE DOES NOT IMPLY DISCRIMINATIVE COLLAPSE

Anonymous authors
Paper under double-blind review

ABSTRACT

Recursive training on synthetic data can lead to *model collapse*, in which learned generative distributions progressively lose diversity or degenerate across training generations. In this paper, we ask whether such generative degeneration necessarily deteriorates the downstream task performance. We distinguish *generative collapse*, meaning collapse of the learned data distribution, from *discriminative collapse*, referring to degradation in performance on downstream tasks such as classification or regression. We theoretically investigate this distinction in a solvable setting: the iterative maximum likelihood estimation (MLE) for the symmetric binary Gaussian mixture model under the discard workflow, where each generation is trained only on synthetic samples from the previously fitted model. In line with a widely known covariance collapse phenomena under the multivariate Gaussian distribution, the fitted covariance matrix also vanishes to zero almost surely, so the trained generative model becomes degenerate under the symmetric binary GMM. Our main finding is that this generative collapse does not necessarily imply collapse of the performance on the downstream classification task. Although the covariance collapses, the fitted mean vector converges almost surely to a non-degenerate random vector that retains directional information about the population signal. Consequently, the regularized plug-in linear discriminant analysis (LDA) classifier has the limiting expected risk strictly smaller than $1/2$, i.e., it is superior to random guessing. Moreover, in the high-SNR regime, its limiting expected risk approaches to the Bayes risk at a quadratic rate keeping the condition number of the ground truth covariance matrix constant. We further show that within the large-sample regime as $n \to \infty$, the terminal mean vector converges weakly to the multivariate t-distribution with degrees of freedom 4, centered at the true mean vector. These results show that generative collapse does not always lead to discriminative collapse.

1 INTRODUCTION

In recent years, synthetic data has gained increasing attention as a common part of model training pipelines. Large language models (LLMs), diffusion models, and other generative models are now used not only to produce final outputs, but also to generate data that may later be used for training. This makes iterative training on synthetic data difficult to avoid. For example, an image search query today may return a mixture of real photographs and AI-generated images. If these images are later scraped and used to train the next generation of image models, then the training set already contains outputs from previous models. Similar feedback loops can occur for text, code, and other forms of data collected from the web. Thus, recursive use of synthetic data is not only a hypothetical training protocol, but an increasingly natural consequence of how data are generated and collected.

A number of recent works have shown that this feedback loop can lead to *model collapse* Shumailov et al. (2023; 2024); Dohmatob et al. (2024a); Seddik et al. (2024). In these works, models trained repeatedly on synthetic data may lose diversity, forget rare events, or converge to overly concentrated distributions. This phenomenon has been studied in language models Shumailov et al. (2024); Guo et al. (2024); Wright et al. (2025), diffusion and other generative models Yoon et al. (2024); Gong et al. (2024), and statistical estimation problems Dohmatob et al. (2024b); Seddik et al. (2024); Dey & Donoho (2024); Amin et al. (2026). The main message is that synthetic data are not, in general,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

a harmless replacement for real data: errors and biases in one training generation can be propagated to later generations and amplified.

Most of this literature studies model collapse as a *generative* phenomenon. That is, the focus is on whether the learned distribution continues to match the original data distribution. For instance, prior works center on loss of tail information, changes in scaling laws, contraction of support, or collapse of variance Dohmatob et al. (2024a;b); Xu et al. (2025). Other work studies when iterative retraining remains stable or consistent Bertrand et al. (2024); Barzilai & Shamir (2025), or proposes ways to reduce collapse by keeping original real data, accumulating synthesized data across training generations, using verification, or choosing appropriate mixing weights Gerstgrasser et al. (2024); Feng et al. (2025); Yi et al. (2026); He et al. (2025); Garg et al. (2026). These results give a detailed picture of when the learned generative model does or does not remain faithful to the original distribution.

Less is known about what happens to downstream prediction. In a flurry of applications, the ultimate goal is not to sample from the learned distribution, but to use the trained model for a downstream task such as classification or regression. This leads to a different question. If iterative training causes the learned data distribution to collapse, does the downstream classifier also collapse? In other words, does generative collapse necessarily imply discriminative collapse?

We distinguish between these two notions:

- **Generative collapse**: degeneration of the learned data distribution itself.
- **Discriminative collapse**: collapse of performance on downstream prediction tasks such as classification or regression.

These two notions are related, but they are not identical. A generative model may fail to represent the full data distribution while still preserving information that is relevant for a particular prediction task. The goal of this paper is to study this distinction in a setting where the training dynamics can be analyzed explicitly. This paper aims to initiate a systematic study of *discriminative collapse* under the iterative maximum likelihood estimation (MLE) within a tractable statistical framework.

We summarize our contributions as follows.

- **Generative and discriminative collapse can separate.** We study recursive MLE under the *discard-workflow* (details provided in the Section 2.1) for the symmetric binary Gaussian mixture model. In Section 3, we showed that the fitted covariance collapses to zero, so the learned generative model becomes degenerate, but the fitted mean converges to a non-degenerate random limit.
- **The limiting classifier remains non-trivial.** In Section 4.2, we analyze the regularized plug-in LDA classifier induced by the recursively fitted model. Its limiting expected risk is strictly below $1/2$, showing that generative collapse does not necessarily reduce downstream classification to random guessing. This has also been supported by the numerical experiments done in Section 6.1 under various structures of the true covariance matrix.
- **High-SNR classification is near optimal.** In Section 4.2, we showed that when the Mahalanobis *signal-to-noise* ratio is large, the limiting expected risk of the regularized plug-in LDA classifier approaches the Bayes risk at a quadratic rate.
- **The residual signal has an explicit limit law.** In Section 5, we showed that in the large-sample regime, the terminal mean vector converges weakly to a multivariate t-distribution with four degrees of freedom, centered at the ground-truth mean. This has also been supported by the numerical experiments done in Section 6.2.

1.1 Related work

Generative model collapse. A recent line of work has shown that recursive training on synthetic data can lead to *model collapse*. This phenomenon has been studied in language models, diffusion models, regression, and parametric estimation Alemohammad et al. (2024); Dohmatob et al. (2024a;b); Seddik et al. (2024); Shumailov et al. (2023; 2024); Yoon et al. (2024). The main message is that synthetic data are not a neutral substitute for real data when they are reused across training generations. Errors in the learned model can be passed to later rounds of training, leading

to loss of tail information, reduced diversity, or collapse of variance. In statistical models, this can often be seen directly: the fitted distribution may become more concentrated even when the original distribution is not degenerate. Several works also study when recursive training is stable, or how collapse can be avoided. For example, Barzilai & Shamir (2025) and Bertrand et al. (2024) prove consistency and stability results for iterative retraining. Other papers propose mitigation strategies, such as keeping real data, accumulating data across generations, or choosing suitable real-synthetic mixing weights Dey & Donoho (2024); Garg et al. (2026); Gerstgrasser et al. (2024); He et al. (2025). These works are close to ours in motivation, but they mainly ask whether the learned generative distribution remains faithful to the original one. We instead allow the generative model to collapse and ask what this means for a downstream classifier.

Discriminative effects of synthetic data. Another line of work studies synthetic data through downstream prediction rather than generative fidelity. Feng et al. Feng et al. (2025) consider binary classification with synthetic labels and a verifier, and show that filtering low-quality synthetic examples can recover Bayes-optimal performance in suitable regimes. Firdoussi et al. Firdoussi et al. (2025) study related questions in high-dimensional models with feature shift and label noise. These results show that the value of synthetic data can be task-dependent: a synthetic distribution may be imperfect, but still useful if its errors do not damage the relevant decision boundary. Other recent papers study verification mechanisms more directly. Verifier-based methods Yi et al. (2026) and self-verification methods Fu et al. (2026) show that filtering synthetic samples can improve recursive training and prevent some forms of collapse. These works are important for understanding how synthetic data can be used safely in prediction pipelines. However, the improvement there comes from an additional selection or correction step. The limiting behavior therefore depends not only on the generator, but also on the verifier and its possible biases.

2 Problem formulation

We now formalize a tractable setting in which the distinction between generative and discriminative collapse can be analyzed exactly. The model is a *symmetric binary Gaussian mixture model (GMM)*, where the same parameters determine both the learned data distribution and the Bayes classifier:

$$(\mathbf{X}, Y) \sim \mathbb{P}_{(\boldsymbol{\mu},\boldsymbol{\Sigma})} \quad \Leftrightarrow \quad Y \sim \mathrm{Unif}\left(\{\pm 1\}\right) \quad \text{and} \quad \mathbf{X}\,|Y \sim \mathcal{N}_d\left(Y\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \tag{2.1}$$

with a model parameter $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta := \mathbb{R}^d \times \mathcal{S}^d_{++}$. In this paper, we assume that the ground-truth mean vector is non-zero, i.e., $\boldsymbol{\mu}_\star \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$.

2.1 The discard-workflow of iterative MLE

Let us focus on the *discard-workflow*, in which each training generation discards all previous data and trains only on fresh synthetic data generated via the model fitted in the preceding generation. This workflow isolates the effect of recursive training on synthesized data: no real data are retained, and no samples from earlier generations are accumulated.

Under the symmetric binary Gaussian mixture model equation 2.1, the maximum likelihood estimation (MLE) update rule has a closed-form formula. Starting from the ground-truth parameter $\left(\widehat{\boldsymbol{\mu}}_0(n), \widehat{\boldsymbol{\Sigma}}_0(n)\right) := (\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$, we draw n synthetic samples

$$O_i^{(t)} := \left(\mathbf{X}_i^{(t)}, Y_i^{(t)}\right) \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\left(\widehat{\boldsymbol{\mu}}_t(n), \widehat{\boldsymbol{\Sigma}}_t(n)\right)}, \quad i \in [n], \tag{2.2}$$

in the t-th training generation. We next compute the MLE based on the synthetic data equation 2.2 as

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_{t+1}(n) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i^{(t)} Y_i^{(t)}, \\ \widehat{\boldsymbol{\Sigma}}_{t+1}(n) &= \frac{1}{n}\sum_{i=1}^{n} \left\{\mathbf{X}_i^{(t)} - Y_i^{(t)}\widehat{\boldsymbol{\mu}}_{t+1}(n)\right\}\left\{\mathbf{X}_i^{(t)} - Y_i^{(t)}\widehat{\boldsymbol{\mu}}_{t+1}(n)\right\}^\top. \end{aligned} \tag{2.3}$$

See Algorithm 1 for the summarization of the resulting recursion.

Algorithm 1 The discard-workflow of iterative MLE under the symmetric binary GMM equation 2.1

Input: A fixed number of synthetic samples $n \in \mathbb{N}$ and the number of training generations $T \in \mathbb{N}$.

1: Set $\left(\widehat{\boldsymbol{\mu}}_0(n), \widehat{\boldsymbol{\Sigma}}_0(n)\right) := (\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta = \mathbb{R}^d \times \mathcal{S}^d_{++}$.

2: **for** $t = 0, 1, 2, \cdots, T-1$ **do**:

3: Sample $\mathbf{O}^{(t)}_{1:n} := \left(O^{(t)}_i := \left(\mathbf{X}^{(t)}_i, Y^{(t)}_i\right) : i \in [n]\right) \sim \mathbb{P}^{\otimes n}_{\left(\widehat{\boldsymbol{\mu}}_t(n), \widehat{\boldsymbol{\Sigma}}_t(n)\right)}$.

4: Train model on the synthetically generated data $\mathbf{O}^{(t)}_{1:n}$ via the following update rule:

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_{t+1}(n) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}^{(t)}_i Y^{(t)}_i, \\ \widehat{\boldsymbol{\Sigma}}_{t+1}(n) &:= \frac{1}{n}\sum_{i=1}^{n} \left\{\mathbf{X}^{(t)}_i - Y^{(t)}_i \widehat{\boldsymbol{\mu}}_{t+1}(n)\right\}\left\{\mathbf{X}^{(t)}_i - Y^{(t)}_i \widehat{\boldsymbol{\mu}}_{t+1}(n)\right\}^\top. \end{aligned} \tag{2.4}$$

5: **end for**

6: **return** the model parameter $\left(\widehat{\boldsymbol{\mu}}_T(n), \widehat{\boldsymbol{\Sigma}}_T(n)\right) \in \Theta$.

2.2 Classification risk and discriminative collapse

The recursion above describes the learned generative model. To evaluate whether its collapse also harms downstream prediction, we next define the classification task and the corresponding risk under the original data-generating distribution.

Definition 2.1 (Classification risk)**.** For any binary classifier $f : \mathbb{R}^d \to \{\pm 1\}$, its *classification risk* under the ground-truth symmetric binary GMM $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$ is defined as

$$\mathcal{R}(f) := \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left\{Y \neq f(\mathbf{X})\right\}.$$

Consider the linear classifier induced by the weight vector $\mathbf{w} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ as

$$f_{\mathbf{w}} : \mathbb{R}^d \to \{\pm 1\}, \quad f_{\mathbf{w}}(\mathbf{x}) := 2 \cdot \mathbb{1}_{\{\mathbf{w}^\top \mathbf{x} \geq 0\}} - 1. \tag{2.5}$$

For notational simplicity, let $\mathbf{w}_\star := \boldsymbol{\Sigma}_\star^{-1}\boldsymbol{\mu}_\star \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ denote the *Bayes direction* and $f_\star := f_{\mathbf{w}_\star} : \mathbb{R}^d \to \{\pm 1\}$ stand for the *Bayes optimal classifier*, i.e.,

$$f_\star = f_{\mathbf{w}_\star} \in \arg\min\left\{\mathcal{R}(f) : f \in \left(\mathbb{R}^d \to \{\pm 1\}\right)\right\}. \tag{2.6}$$

We are now in position to characterize the risk for the linear classifier $f_{\mathbf{w}} : \mathbb{R}^d \to \{\pm 1\}$ under the ground-truth symmetric binary GMM $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$. See Appendix B.1 for the proof of Proposition 2.1.

Proposition 2.1. *For any weight vector $\mathbf{w} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$, the risk of the binary classifier $f_{\mathbf{w}} : \mathbb{R}^d \to \{\pm 1\}$ under the ground-truth symmetric binary GMM $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$ is given by*

$$\mathcal{R}\left(f_{\mathbf{w}}\right) = \Phi\left(-\frac{\mathbf{w}^\top \boldsymbol{\mu}_\star}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w}}}\right), \tag{2.7}$$

where $\Phi : \mathbb{R} \to [0, 1]$ stands for the cumulative distribution function (CDF) of the standard normal distribution $\mathcal{N}(0, 1)$, i.e., $\Phi(x) := \int_{-\infty}^{x} \phi(t)\mathrm{d}t$ and $\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ for $x \in \mathbb{R}$.

3 Generative collapse under the discard-workflow

This section takes a step forward to understand the core behavior of the learning dynamics produced by the discard-workflow of iterative MLE. The covariance collapse component from Proposition 3.2 is consistent with standard variance-collapse mechanisms under the discard-workflow of recursive model training. One key takeaway message is that this covariance collapse does not force the fitted mean vector to vanish: instead, it converges to a non-degenerate d-dimensional random vector.

3.1 EXACT TRANSITION LAW OF THE LEARNING DYNAMICS

We first introduce the *canonical filtration* $\mathcal{F}_\star(n) := (\mathcal{F}_t(n) : t \in \mathbb{Z}_+)$ on $\mathbb{R}^d \times \{\pm 1\}$, where

$$\mathcal{F}_0(n) := \left\{\varnothing, \mathbb{R}^d \times \{\pm 1\}\right\} \quad \text{and} \quad \mathcal{F}_t(n) := \sigma\left(\mathbf{O}_{1:n}^{(0)}, \mathbf{O}_{1:n}^{(1)}, \cdots, \mathbf{O}_{1:n}^{(t-1)}\right), \quad t \geq 1, \tag{3.1}$$

denotes the σ-field generated from the sample trajectory of the discard-workflow of iterative MLE (Algorithm 1). It can be easily realized from the update rule equation 2.4 that the stochastic process

$$\left(\left(\widehat{\boldsymbol{\mu}}_t(n), \widehat{\boldsymbol{\Sigma}}_t(n)\right) : t \in \mathbb{Z}_+\right)$$

forms a discrete-time Markov chain (DTMC) with respect to the canonical filtration $\mathcal{F}_\star(n)$. We can explicitly characterize its transition probability law through the following proposition.

Proposition 3.1. *Given any ground-truth model parameter $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$, it follows for any fixed sample size $n \geq d+1$ and $T \in \mathbb{Z}_+$ that*

(i) $\widehat{\boldsymbol{\mu}}_{T+1}(n) | \, \mathcal{F}_T(n) \sim \mathcal{N}_d\left(\widehat{\boldsymbol{\mu}}_T(n), \frac{1}{n}\widehat{\boldsymbol{\Sigma}}_T(n)\right)$.

(ii) $n \cdot \widehat{\boldsymbol{\Sigma}}_{T+1}(n) \Big| \, \mathcal{F}_T(n) \;\sim\; \mathbb{W}_d\left(n-1, \widehat{\boldsymbol{\Sigma}}_T(n)\right)$, *where $\mathbb{W}_p(k, \mathbf{V})$ is the p-dimensional Wishart distribution with degrees of freedom $k \in \mathbb{N}$ and scale matrix $\mathbf{V} \in \mathcal{S}_{++}^p$.*

Consequently, we obtain

$$\widehat{\boldsymbol{\mu}}_{T+1}(n) | \, \mathcal{V}_T(n) \sim \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{1}{n}\sum_{t=0}^{T} \widehat{\boldsymbol{\Sigma}}_t(n)\right), \tag{3.2}$$

where $\mathcal{V}_T(n) = \sigma\left(\widehat{\boldsymbol{\Sigma}}_t(n) : 0 \leq t \leq T\right)$ and the conditional contraction identity for the sequence $\left\{\widehat{\boldsymbol{\Sigma}}_t(n) : t \in \mathbb{Z}_+\right\}$:

$$\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\Sigma}}_{T+1}(n)\right] = \left(1 - \frac{1}{n}\right)\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\Sigma}}_T(n)\right], \tag{3.3}$$

thereby concluding that $\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\Sigma}}_T(n)\right] = \left(1 - \frac{1}{n}\right)^T \boldsymbol{\Sigma}_\star \to \mathbf{O}_{d\times d}$ as $T \to \infty$.

Proposition 3.1 separates the two components of the update rule equation 2.4 in Algorithm 1. The covariance matrix evolves multiplicatively through the Wishart updates, while the mean vector evolves additively with conditional noise proportional to the covariance matrix at the current stage. We refer the readers to Appendix C.1 for the proof of Proposition 3.1.

3.2 COLLAPSE OF THE COVARIANCE MATRIX AND THE CONVERGENCE OF THE MEAN VECTOR

Several recent studies have proved that the discard-workflow of iterative MLE (Algorithm 2) leads to variance collapse in the multivariate Gaussian model Shumailov et al. (2024); Alemohammad et al. (2024); Bertrand et al. (2024). We formalize the analogous result for the symmetric binary GMM equation 2.1 through the following proposition.

Proposition 3.2 (Covariance collapse)**.** *For any ground-truth parameter $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$ with $\boldsymbol{\mu}_\star \neq \mathbf{0}_d$ and any fixed sample size $n \geq d+1$, we have $\widehat{\boldsymbol{\Sigma}}_T(n) \to \mathbf{O}_{d\times d}$ as $T \to \infty$ $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost surely.*

The proof of Proposition 3.2 is deferred to Appendix C.2. Proposition 3.2 basically demonstrates *generative collapse* of the trained symmetric binary GMM under the discard-workflow of iterative MLE (Algorithm 2).

However, the asymptotic behavior for the fitted mean vectors is different. Since $\{\widehat{\boldsymbol{\mu}}_t(n) : t \in \mathbb{Z}_+\}$ is an $\mathcal{F}_\star(n)$-martingale and the covariance matrix of its increment is proportional to the fitted covariance matrix at the current step, collapse of $\widehat{\boldsymbol{\Sigma}}_T(n)$ anneals the future noise injected into the

update rule of the mean vector. This leads to the convergence of the fitted mean vector to some non-degenerate terminal mean vector, which can be formally stated as Theorem 3.1. See Appendix C.3 for its proof.

Theorem 3.1 (Convergence of the fitted mean vector)**.** *Given any ground-truth model parameter* $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$ *with* $\boldsymbol{\mu}_\star \neq \mathbf{0}_d$ *and any fixed sample size* $n \geq d+1$*, there exists some* d*-dimensional random vector* $\widehat{\boldsymbol{\mu}}_\infty(n) \in L^2\left(\mathbb{R}^d \times \{\pm 1\}, \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\right)$ *such that*

$$\widehat{\boldsymbol{\mu}}_T(n) \overset{T\to\infty}{\longrightarrow} \widehat{\boldsymbol{\mu}}_\infty(n) \quad \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\text{-almost surely and in } L^2. \tag{3.4}$$

Moreover, we have

$$\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_\infty(n)\right] = \boldsymbol{\mu}_\star \quad \text{and} \quad \mathrm{Cov}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_\infty(n)\right] = \boldsymbol{\Sigma}_\star. \tag{3.5}$$

4 Does generative collapse lead to discriminative collapse?

Equipped with the generative collapse results established in Section 3 in hand, we now ask whether this generative collapse forces the performance for the downstream classification task to collapse. In order to ask this question, we study the *plug-in linear discriminant analysis (LDA) classifier* induced by the discard-workflow of iterative MLE (Algorithm 1). Since $\widehat{\boldsymbol{\Sigma}}_T(n)$ vanishes to zero as $T \to \infty$ by virtue of Proposition 3.2, the *unregularized plug-in LDA direction* $\widehat{\mathbf{w}}_T(n) := \widehat{\boldsymbol{\Sigma}}_T^{-1}(n)\widehat{\boldsymbol{\mu}}_T(n)$ turns out to be asymptotically ill-conditioned. We therefore consider the *regularized plug-in LDA classifier* instead and aim to understand the limiting behavior of its expected risk.

4.1 The regularized plug-in LDA classifier

Since Proposition 3.2 shows that $\widehat{\boldsymbol{\Sigma}}_T(n)$ becomes singular as $T \to \infty$, let us introduce the regularized plug-in LDA classifier:

$$\widehat{\mathbf{w}}_T^\delta(n) := \left\{\widehat{\boldsymbol{\Sigma}}_T(n) + \delta \mathbf{I}_d\right\}^{-1} \widehat{\boldsymbol{\mu}}_T(n) \quad \text{and} \quad \hat{f}_T^\delta(n) := f_{\widehat{\mathbf{w}}_T^\delta(n)} : \mathbb{R}^d \to \{\pm 1\} \tag{4.1}$$

for some regularization parameter $\delta \in (0, +\infty)$. Then, Proposition 3.2 and Theorem 3.1 imply that for any fixed sample size $n \geq d+1$, the following $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost sure convergence holds:

$$\mathcal{R}\left\{\hat{f}_T^\delta(n)\right\} \overset{T\to\infty}{\longrightarrow} \Phi\left(-\frac{\boldsymbol{\mu}_\star^\top \widehat{\boldsymbol{\mu}}_\infty(n)}{\sqrt{\left\{\widehat{\boldsymbol{\mu}}_\infty(n)\right\}^\top \boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\mu}}_\infty(n)}}\right) =: \mathcal{R}_\infty(n). \tag{4.2}$$

Therefore, one needs to take a closer inspection on the random variable $\mathcal{R}_\infty(n)$ in order to study the limiting behavior of the risk for the regularized plug-in LDA classifier equation 4.1 as $T \to \infty$, which turns out to be independent of the choice of the regularization parameter $\delta \in (0, +\infty)$. In particular, we focus on the expectation of $\mathcal{R}_\infty(n)$ to understand the asymptotic behavior of the expected risk for the regularized plug-in LDA classifier equation 4.1 as $T \to \infty$, from the fact that

$$\lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}\left\{\hat{f}_T^\delta(n)\right\}\right] = \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right], \tag{4.3}$$

which follows by virtue of the bounded convergence theorem.

4.2 The regularized plug-in LDA classifier asymptotically beats random guessing

Our next result elucidates that the limiting expected risk for the regularized plug-in LDA classifier $\hat{f}_T^\delta(n) : \mathbb{R}^d \to \{\pm 1\}$ is strictly less than $1/2$, thereby claiming that the discard-workflow of iterative MLE with unknown covariance matrix turns out to be better than random guessing. See Appendix D.2 for the proof of Theorem 4.1.

Theorem 4.1. *For any given ground-truth model parameter* $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$ *with* $\boldsymbol{\mu}_\star \neq \mathbf{0}_d$*, any fixed sample size* $n \geq d+1$*, and any regularization parameter* $\delta \in (0, +\infty)$*, we have*

$$\lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}\left\{\hat{f}_T^\delta(n)\right\}\right] = \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right] < \frac{1}{2}. \tag{4.4}$$

Theorem 4.1 argues that even though the learned symmetric binary GMM degenerates, the regularized plug-in LDA direction retains a non-trivial alignment with the Bayes direction asymptotically. While Theorem 4.1 only shows that the regularized plug-in LDA classifier equation 4.1 is superior than random guessing as $T \to \infty$, we next establish that it achieves a near-optimal performance within the high-SNR regimes, where the *Mahalanobis signal-to-noise ratio (SNR)* is given by $\Delta_\star := \sqrt{\boldsymbol{\mu}_\star^\top \boldsymbol{\Sigma}_\star^{-1} \boldsymbol{\mu}_\star}$.

Theorem 4.2. *For any fixed sample size $n \geq d+1$ and any regularization parameter $\delta \in (0, +\infty)$, it holds that*

$$\lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\mathcal{R}\left\{ \hat{f}_T^\delta(n) \right\} \right] - \mathcal{R}_\star = \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\mathcal{R}_\infty(n) \right] - \mathcal{R}_\star = \mathcal{O}\left(\frac{4d\kappa(\boldsymbol{\Sigma}_\star)^2}{\Delta_\star^2} \right), \tag{4.5}$$

where $\kappa(\boldsymbol{\Sigma}_\star) = \lambda_{\max}(\boldsymbol{\Sigma}_\star)/\lambda_{\min}(\boldsymbol{\Sigma}_\star)$. In particular, along any family of parameters with $\kappa(\boldsymbol{\Sigma}_\star) = \mathcal{O}(1)$, we have

$$\lim_{\Delta_\star \to +\infty} \left\{ \lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\mathcal{R}\left\{ \hat{f}_T^\delta(n) \right\} \right] - \mathcal{R}_\star \right\} = 0. \tag{4.6}$$

The proof of Theorem 4.2 can be found in Appendix D.3. To sum up, when the population classes are well-separated, the decision boundary retained after generative collapse is sufficient for near-Bayes classification.

4.3 Is knowing the covariance matrix helpful?

The preceding results may seem counterintuitive: the classifier avoids random guessing precisely in the regime where the covariance estimate collapses. In order to clarify the mechanism, we compare against the oracle variant of the discard-workflow of iterative MLE under the symmetric binary GMM equation 2.1 in which the learner is given with the ground-truth covariance matrix $\boldsymbol{\Sigma}_\star \in \mathcal{S}_{++}^d$ at every training generation and only updates the mean vector. In this case, the corresponding update rule is given by

$$\mathbf{O}_{1:n}^{(t)} = \left(O_i^{(t)} = \left(\mathbf{X}_i^{(t)}, Y_i^{(t)} \right) : i \in [n] \right) \sim \mathbb{P}_{(\widehat{\boldsymbol{\mu}}_t^\star(n), \boldsymbol{\Sigma}_\star)}^{\otimes n}, \quad \widehat{\boldsymbol{\mu}}_{t+1}^\star(n) := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^{(t)} Y_i^{(t)}. \tag{4.7}$$

Let $\widehat{\mathbf{w}}_T^\star(n) := \boldsymbol{\Sigma}_\star^{-1} \widehat{\boldsymbol{\mu}}_T^\star(n)$ and $\widehat{f}_T^\star(n) := f_{\widehat{\mathbf{w}}_T^\star(n)} : \mathbb{R}^d \to \{\pm 1\}$ denote the plug-in LDA direction and the induced linear classifier, respectively, at the T-th training generation of the discard-workflow of iterative MLE when $\boldsymbol{\Sigma}_\star \in \mathcal{S}_{++}^d$ is revealed to us. Unlike the original update rule equation 2.4, the noise injected into the mean vector update does not decay over training generations. The fitted mean vector therefore continues to diffuse, and the induced plug-in LDA classifier $\widehat{f}_T^\star(n) : \mathbb{R}^d \to \{\pm 1\}$ becomes asymptotically no better than random guessing.

Theorem 4.3. *Under the discard-workflow of iterative MLE under the symmetric binary GMM equation 2.1 with known covariance matrix, we have for any fixed sample size $n \geq d+1$ that*

$$\lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\mathcal{R}\left\{ \widehat{f}_T^\star(n) \right\} \right] = \frac{1}{2}. \tag{4.8}$$

This comparison highlights the key rationale behind the discriminative non-collapse for the discard-workflow of iterative MLE with unknown $\boldsymbol{\Sigma}_\star \in \mathcal{S}_{++}^d$. The collapsing covariance matrix is harmful in a generative manner, but it also anneals the future noise injected into the update rule for the fitted mean vector. As a consequence, the fitted mean vector converges to a non-degenerated d-dimensional random vector rather than diffusing indefinitely. The regularized plug-in LDA classifier equation 4.1 then extracts the surviving signal from the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$.

5 Large-sample law of the residual signal after generative collapse

Our discussions so far elucidate that the downstream task performance is governed by the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$. We now characterize the residual signal carried by the terminal vector under the

large-sample regime. The limit is heavy-tailed: after infinitely many recursive training generations, and then as $n \to \infty$, the terminal mean vector converges to a multivariate t-distribution centered at the true mean vector. We recall the definition of the multivariate t-distribution: Let $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \in \mathbb{R}^d \times \mathcal{S}_{++}^d \times (0, +\infty)$. The *$d$-dimensional t-distribution with a location parameter $\boldsymbol{\mu} \in \mathbb{R}^d$, a scale parameter $\boldsymbol{\Sigma} \in \mathcal{S}_{++}^d$, and the degrees of freedom $\nu \in (0, +\infty)$* stands for the probability distribution of a d-dimensional random vector $\mathbf{X}$, which is often denoted by $t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where

$$\mathbf{X} = \boldsymbol{\mu} + \frac{\mathbf{Z}}{\sqrt{\frac{G}{\nu}}}, \quad \mathbf{Z} \sim \mathcal{N}_d(\mathbf{0}_d, \boldsymbol{\Sigma}) \perp\!\!\!\perp G \sim \Gamma\left(\frac{\nu}{2}, 2\right). \tag{5.1}$$

Theorem 5.1 (Large-sample law for the terminal mean vector)**.** *Given any ground-truth parameter $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$ with $\boldsymbol{\mu}_\star \neq \mathbf{0}_d$, the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ produced by the discard-workflow of iterative MLE under the symmetric binary GMM equation 2.1 (Algorithm 1) satisfies*

$$\widehat{\boldsymbol{\mu}}_\infty(n) \Rightarrow t_d\left(\boldsymbol{\mu}_\star, \frac{1}{2}\boldsymbol{\Sigma}_\star, 4\right) \quad \text{as } n \to \infty. \tag{5.2}$$

Let us briefly discuss the key ideas for the proof of Theorem 5.1. We aim to compute the limit of the (unconditional) characteristic function of the d-dimensional random vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ as $n \to \infty$. Since the characteristic function of $\widehat{\boldsymbol{\mu}}_\infty(n)$ can be expressed as the Laplace transform of the random variable $\frac{\mathfrak{R}(\mathbf{v};n)}{n}$, where the *random perpetuity* $\mathfrak{R}(\mathbf{v};n)$ is defined in equation E.12, we focus on the weak convergence of $\widehat{\boldsymbol{\mu}}_\infty(n)$ as $n \to \infty$ by virtue of the continuous mapping theorem. Specifically, we establish that $\frac{\mathfrak{R}(\mathbf{v};n)}{n}$ weakly converges to the inverse Gamma distribution with parameter $(2, 1)$ (Theorems E.1 and E.2) using the continuous-time martingale functional central limit theorem (Lemma A.9) together with various advanced probability theory tools. As it turns out, the limit of the characteristic function of $\widehat{\boldsymbol{\mu}}_\infty(n)$ coincides with the characteristic function of $t_d\left(\boldsymbol{\mu}_\star, \frac{1}{2}\boldsymbol{\Sigma}_\star, 4\right)$ as $n \to \infty$, concluding the proof of Theorem 5.1. We defer the full technical details to Appendix E.1.

6 Numerical experiments

In this section, we complement our theory with two numerical simulation results. Both run the *exact* transition probability law of the DTMC governed by the discard-workflow of iterative MLE (Proposition 3.1): each generation is a single batched $\widehat{\boldsymbol{\mu}}_{t+1}(n) \,\big|\, \left(\widehat{\boldsymbol{\mu}}_t(n), \widehat{\boldsymbol{\Sigma}}_t(n)\right) \sim \mathcal{N}_d\left(\widehat{\boldsymbol{\mu}}_t(n), \frac{1}{n}\widehat{\boldsymbol{\Sigma}}_t(n)\right)$ together with $\widehat{\boldsymbol{\Sigma}}_{t+1}(n) \,\big|\, \widehat{\boldsymbol{\Sigma}}_t(n) \sim \frac{1}{n}\mathbb{W}_d\left(n-1, \widehat{\boldsymbol{\Sigma}}_t(n)\right)$, with no re-sampling of synthetic data. The classification risks are evaluated using the closed-form formula provided in Proposition 2.1; no test dataset is drawn.

6.1 Discriminative non-collapse

Figure 1 reports the expected classification risk as a function of the training generation T, comparing the regularized plug-in LDA classifier $\hat{f}_T^\delta(n) : \mathbb{R}^d \to \{\pm 1\}$ defined in equation 4.1 against the oracle plug-in LDA classifier $\hat{f}_T^\star(n) : \mathbb{R}^d \to \{\pm 1\}$ utilizing the knowledge of the ground-truth covariance matrix $\boldsymbol{\Sigma}_\star \in \mathcal{S}_{++}^d$. We consider the parameters $(d, n, T) = (10, 30, 5000)$ together with the regularization parameter $\delta = 10^{-6}$, and average over 200 independent DTMCs. In order to isolate the influence of the structure of the covariance matrix, $\boldsymbol{\mu}_\star \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ is re-scaled in each configuration so that the Mahalanobis SNR is fixed as a constant value of $\Delta_\star = \sqrt{\boldsymbol{\mu}_\star^\top \boldsymbol{\Sigma}_\star^{-1} \boldsymbol{\mu}_\star} = 1.5$; consequently, the Bayes risk $\mathcal{R}_\star = \Phi(-\Delta_\star) \approx 0.0668$ is identical across the three panels. We consider the following three different structures: (i) an isotropic covariance structure $\boldsymbol{\Sigma}_\star = \mathbf{I}_d$; (ii) an anisotropic covariance structure $\boldsymbol{\Sigma}_\star = \mathrm{diag}(1, 2, \cdots, d)$; and (iii) an AR(1)-correlated structure $[\boldsymbol{\Sigma}_\star]_{ij} = 0.7^{|i-j|}$ for each $(i, j) \in [d] \times [d]$.

In every panel, the curve of the expected risk for the oracle plug-in LDA classifier $\hat{f}_T^\star(n)$ drifts toward the risk of the random guessing $1/2$, in agreement with Theorem 4.3. The curve of the expected risk for the regularized plug-in LDA classifier $\hat{f}_T^\delta(n)$, in contrast, plateaus strictly below

$1/2$ in all three configurations. The plateau value depends on the covariance structure – it sits much closer to $1/2$ for the heavy-spectrum anisotropic case for the covariance matrix compared to the isotropic and AR(1) cases.

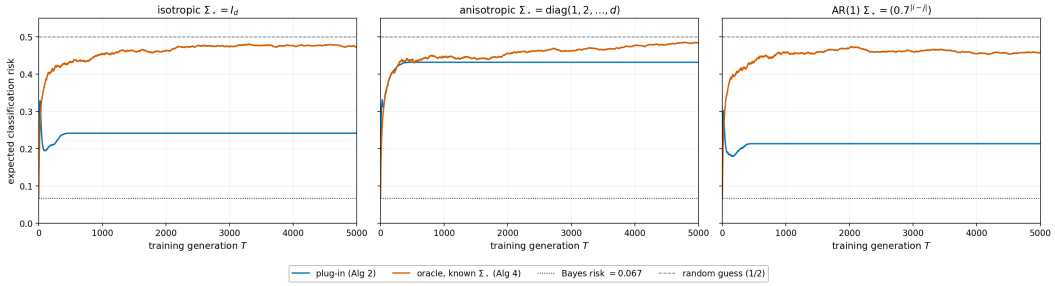


Figure 1: Plots for the expected classification risk across the training generations under the discard-workflow for three configurations of $\mathbf{\Sigma}_\star \in \mathcal{S}^d_{++}$. The expected risk of the regularized plug-in LDA classifier $\hat{f}^{\delta}_T(n)$ plateaus strictly below $1/2$, while that of the oracle plug-in LDA classifier $\hat{f}^{\star}_T(n)$ drifts toward $1/2$.

6.2 Distributional characterization of the terminal mean vector

Figure 2 corroborates the exact distributional characterization of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ from Theorem 5.1, which proclaims that $\widehat{\boldsymbol{\mu}}_\infty(n) \Rightarrow t_d\left(\boldsymbol{\mu}_\star, \mathbf{\Sigma}_\star/2, 4\right)$ as $n \to \infty$. Take $d = 2$, $\boldsymbol{\mu}_\star = (1, 0)^\top$, $\mathbf{\Sigma}_\star = \mathbf{I}_2$, and run $N = 20{,}000$ independent DTMCs for $T = 10n$ training generations, which is sufficient that the residual covariance $(1 - 1/n)^T < e^{-10}$ is numerically indistinguishable from zero, across $n \in \{10, 50, 200, 1000\}$.

The top row overlays the empirical histogram of the first marginal of $\widehat{\boldsymbol{\mu}}_T(n) - \boldsymbol{\mu}_\star$ with the density function of $t_1\left(0, 1/2, 4\right)$ as predicated by theory and the density function of the normal distribution with matching standard deviation. Even for the case of $n = 10$, the heavy-tails of $t_1\left(0, 1/2, 4\right)$ are visibly the correct fit and those of the Gaussian distributon is plainly inadequate. The bottom row displays the corresponding QQ plots against the $t_1\left(0, 1/2, 4\right)$-quantiles; the empirical points hug the diagonal across the entire range of quantiles, and the agreement tightens as the sample size n grows.

7 Discussion

This paper studies the *discard-workflow* of iterative MLE for the symmetric binary Gaussian mixture model. In this setting, the fitted covariance matrix converges to zero, and hence the learned generative distribution becomes degenerate. In contrast, the fitted mean vector converges to a non-degenerate d-dimensional random vector, rather than being forced to vanish together with the covariance matrix. This distinction is central to our result: although the fitted generative model collapses, the recursively trained model can still retain information that is useful for classification. The covariance collapse makes the learned generative distribution degenerate, but it also reduces the conditional noise in later updates of the fitted mean vector. As a result, the mean vector stabilizes instead of diffusing indefinitely. The regularized plug-in LDA classifier built from the limiting mean vector therefore has limiting expected risk strictly below $1/2$. Moreover, when the Mahalanobis *signal-to-noise* ratio is large, this limiting risk approaches the Bayes risk. Taken together, these results show that *generative collapse* and *discriminative collapse* are distinct phenomena. Even when recursive training drives the learned data distribution to a degenerate limit, the information relevant for a downstream classifier may still survive. In the symmetric binary GMM, this surviving information is carried by the limiting fitted mean vector, which remains non-degenerate and retains alignment with the population signal. Consequently, the collapse of a *generative model* does not automatically determine the fate of downstream prediction.

AI USE STATEMENT

We used AI-assisted tools (LLMs) to polish the writing and improve the clarity of the manuscript. The authors take responsibility for the final content, including the theoretical and numerical results.

ETHICS STATEMENT

This work consists of theoretical investigations under stylized statistical models and supporting numerical simulation results. It does not involve human participants, personal data, or the deployment of learning systems.

REPRODUCIBILITY STATEMENT

The problem formulation, underlying models, and assumptions are provided in Section 2. In particular, the detailed description of the recursive training procedure of our interest is described in Algorithm 1. The detailed proofs for the theoretical results and supporting lemmas are provided in the appendix. The experimental setup and numerical simulation results can be found in Section 6.

REFERENCES

Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard Baraniuk. Self-consuming generative models go MAD. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ShjMHfmPs0.

Kareem Amin, Alex Bie, Weiwei Kong, Umar Syed, and Sergei Vassilvitskii. Learning from synthetic data: Limitations of ERM. In *37th International Conference on Algorithmic Learning Theory*, 2026. URL https://openreview.net/forum?id=6xnfm7s6af.

Daniel Barzilai and Ohad Shamir. When models don't collapse: On the consistency of iterative MLE. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=DrBwhFUqCY.

Quentin Bertrand, Joey Bose, Alexandre Duplessis, Marco Jiralerspong, and Gauthier Gidel. On the stability of iterative retraining of generative models on their own data. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JORAfH2xFd.

Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential, B: Theory of Martingales*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Company, Amsterdam, 1982. ISBN 9780444865267.

Apratim Dey and David Donoho. Universality of the $\pi^2/6$ pathway in avoiding model collapse. *arXiv preprint arXiv:2410.22812*, 2024.

Elvis Dohmatob, Yunzhen Feng, Pu Yang, Francois Charton, and Julia Kempe. A tale of tails: Model collapse as a change of scaling laws. In *Forty-first International Conference on Machine Learning*, 2024a. URL https://openreview.net/forum?id=KVvku47shW.

Elvis Dohmatob, Yuzhen Feng, and Julia Kempe. Model collapse demystified: The case of regression. *Advances in Neural Information Processing Systems*, 37:46979–47013, 2024b.

R. M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2002.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Yunzhen Feng, Elvis Dohmatob, Pu Yang, Francois Charton, and Julia Kempe. Beyond model collapse: Scaling up with synthesized data requires verification. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=MQXrTMonT1.

Aymane El Firdoussi, Mohamed El Amine Seddik, Soufiane Hayou, Reda ALAMI, Ahmed Alzubaidi, and Hakim Hacid. Maximizing the potential of synthetic data: Insights from random matrix theory. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=I9Dsq0cVo9.

Shi Fu, Yingjie Wang, Yuzhu Chen, Li Shen, and Dacheng Tao. Self-verification provably prevents model collapse in recursive synthetic training. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=X5Hk8aMs6w.

Anvit Garg, Sohom Bhattacharya, and Pragya Sur. Preventing model collapse under overparametrization: Optimal mixing ratios for interpolation learning and ridge regression. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=9RdhTvYbX0.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Tomasz Korbak, Henry Sleight, Rajashree Agrawal, John Hughes, Dhruv Bhandarkar Pai, Andrey Gromov, Dan Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=5B2K4LRgmz.

Yanxiang Gong, Zhiwei Xie, Mei Xie, and Xin Ma. Testing generated distributions in gans to penalize mode collapse. In *International Conference on Artificial Intelligence and Statistics*, pp. 442–450. PMLR, 2024.

Yanzhu Guo, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel. The curious decline of linguistic diversity: Training language models on synthetic text. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 3589–3604, 2024.

Hengzhi He, Shirong Xu, and Guang Cheng. Golden ratio weighting prevents model collapse. *arXiv preprint arXiv:2502.18049*, 2025.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, 2 edition, 1991. doi: 10.1007/978-1-4612-0949-2.

Rolando Rebolledo. Central limit theorems for local martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 51(3):269–286, 1980. doi: 10.1007/BF00587353.

Mohamed El Amine Seddik, Suei-Wen Chen, Soufiane Hayou, Pierre Youssef, and Merouane Abdelkader DEBBAH. How bad is training on synthetic data? a statistical analysis of language model collapse. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=t3z6UlV09o.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. The curse of recursion: Training on generated data makes models forget. *arXiv preprint arXiv:2305.17493*, 2023.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. Ai models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, 2024.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, NY, 1996. ISBN 978-1-4757-2545-2. doi: 10.1007/978-1-4757-2545-2.

Ward Whitt. Some useful functions for functional limit theorems. *Mathematics of Operations Research*, 5(1):67–85, 1980. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/3689395.

Chi Song Wong, Joe Masaro, and Tonghui Wang. Multivariate versions of cochran's theorems. *Journal of multivariate analysis*, 39(1):154–174, 1991.

Dustin Wright, Sarah Masud, Jared Moore, Srishti Yadav, Maria Antoniak, Peter Ebert Christensen, Chan Young Park, and Isabelle Augenstein. Epistemic diversity and knowledge collapse in large language models. *arXiv preprint arXiv:2510.04226*, 2025.

Shirong Xu, Hengzhi He, and Guang Cheng. A probabilistic perspective on model collapse. *arXiv preprint arXiv:2505.13947*, 2025.

Bingji Yi, Qiyuan Liu, Yuwei Cheng, and Haifeng Xu. Escaping model collapse via synthetic data verification: Near-term improvements and long-term convergence. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=yfk6c39omW.

Youngseok Yoon, Dainong Hu, Iain Weissburg, Yao Qin, and Haewon Jeong. Model collapse in the self-consuming chain of diffusion finetuning: A novel perspective from quantitative trait modeling. *arXiv preprint arXiv:2407.17493*, 2024.

CONTENTS

1 Introduction **1**

1.1 Related work 2

2 Problem formulation **3**

2.1 The discard-workflow of iterative MLE 3

2.2 Classification risk and discriminative collapse 4

3 Generative collapse under the discard-workflow **4**

3.1 Exact transition law of the learning dynamics 5

3.2 Collapse of the covariance matrix and the convergence of the mean vector 5

4 Does generative collapse lead to discriminative collapse? **6**

4.1 The regularized plug-in LDA classifier 6

4.2 The regularized plug-in LDA classifier asymptotically beats random guessing . . . 6

4.3 Is knowing the covariance matrix helpful? 7

5 Large-sample law of the residual signal after generative collapse **7**

6 Numerical experiments **8**

6.1 Discriminative non-collapse 8

6.2 Distributional characterization of the terminal mean vector 9

7 Discussion **9**

A Preliminaries on probability theory and stochastic processes **15**

A.1 Skorokhod space and its metrization 15

A.2 Weak convergence 15

A.3 Basics of local martingales 16

A.4 A few preliminary facts 18

B Omitted details and proofs for Section 2 **19**

B.1 Proof of Proposition 2.1 19

B.2 Iterative model training via maximum likelihood estimation (MLE) 19

C Omitted details and proofs for Section 3 **20**

C.1 Proof of Proposition 3.1 20

C.2 Proof of Proposition 3.2 22

C.3 Proof of Theorem 3.1 22

D Omitted details and proofs for Section 4 **24**

D.1 Several key lemmas 24

D.2 Proof of Theorem 4.1 24
D.3 Proof of Theorem 4.2 26
D.4 Proof of Theorem 4.3 28
D.5 Proof of Lemma D.1 29
D.6 Proof of Lemma D.2 32
D.7 Proof of Lemma D.3 32

E Omitted details and proofs for Section 5 **34**
E.1 Proof of Theorem 5.1 34
E.2 Proof of Theorem E.1 37
E.3 Proof of Theorem E.2 47

F Omitted details for Section 6 **49**

A PRELIMINARIES ON PROBABILITY THEORY AND STOCHASTIC PROCESSES

This section is devoted to review a few preliminaries on probability theory and stochastic processes that play a critical role for proving the main results so that our paper is self-contained.

A.1 SKOROKHOD SPACE AND ITS METRIZATION

Definition A.1 (Càdlàg function)**.** Let $E \subseteq \mathbb{R}$ be a non-empty interval on the real line and $(\mathcal{M}, \rho)$ be a metric space. A function $f : E \to \mathcal{M}$ is *càdlàg* if for every $t \in E$, it is right-continuous at $t \in E$ and its left-limit $f(t-) := \lim_{s \uparrow t^-} f(s)$ exists in $\mathcal{M}$. The set of all càdlàg functions from E to $\mathcal{M}$ is often denoted by $\mathbb{D}(E; \mathcal{M})$ and is called the *$\mathcal{M}$-valued Skorokhod space defined on E*.

If we denote $\mathbb{C}(E; \mathcal{M})$ by the set of all continuous functions from E to $\mathcal{M}$, then it is evident that $\mathbb{C}(E; \mathcal{M}) \subseteq \mathbb{D}(E; \mathcal{M})$. Throughout this paper, let us simply denote by $\mathbb{D}(E) := \mathbb{D}(E; \mathbb{R})$ and $\mathbb{C}(E) := \mathbb{C}(E; \mathbb{R})$. We shall henceforth center our discussion on the $\mathbb{R}$-valued Skorokhod space defined on E.

Definition A.2 (Skorokhod metric on $\mathbb{D}(E)$)**.** Consider the set of all strictly increasing, continuous bijections from $E \subseteq \mathbb{R}$ to itself (i.e., *wiggles in time*) $\Lambda(E)$, and let

$$\|f\|_E := \sup\{|f(t)| : t \in E\}$$

denote the uniform norm of real-valued functions defined on E. The *Skorokhod metric* σ on $\mathbb{D}(E)$ is defined as

$$\sigma(f, g) := \inf\{\max\{\|\lambda - I_E\|_E, \|f - g \circ \lambda\|_E\} : \lambda \in \Lambda(E)\}, \tag{A.1}$$

where $I_E : E \to E$ is the identity function. The topology on $\mathbb{D}(E)$ induced by the Skorokhod metric σ is referred to as the *Skorokhod topology* on $\mathbb{D}(E)$.

A.2 WEAK CONVERGENCE

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space and $\mathcal{F}_\star := (\mathcal{F}(t) : t \in \mathbb{R}_+)$ be a filtration defined on $(\Omega, \mathcal{F})$. We define for any $t \in \mathbb{R}_+$:

$$\begin{aligned} \mathcal{F}(t+) &:= \bigcap_{\varepsilon \in (0, +\infty)} \mathcal{F}(t + \varepsilon), \\ \mathcal{F}(t-) &:= \sigma\left\{ \bigcup_{s \in [0,t)} \mathcal{F}(s) \right\} \quad \text{and} \quad \mathcal{F}(0-) := \mathcal{F}(0). \end{aligned} \tag{A.2}$$

Definition A.3 (Weak convergence)**.** Let $\mathbb{S}$ be any topological space and $\mathcal{B}(\mathbb{S})$ denote its Borel σ-field. We say that a sequence $\{\mu_n \in \Delta(\mathbb{S}) : n \in \mathbb{N}\}$ of probability measures on $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$ *converges weakly* to $\mu \in \Delta(\mathbb{S})$ if

$$\lim_{n \to \infty} \int_{\mathbb{S}} f(x) \mu_n(\mathrm{d}x) = \int_{\mathbb{S}} f(x) \mu(\mathrm{d}x)$$

for every bounded continuous function $f : \mathbb{S} \to \mathbb{R}$, and one often writes it as $\mu_n \Rightarrow \mu$ or $\mu_n \xrightarrow{d} \mu$ as $n \to \infty$. Also, we say that a sequence $\{X_n \in (\Omega \to \mathbb{S}) : n \in \mathbb{N}\}$ of $\mathbb{S}$-valued random elements *converges weakly* to an $\mathbb{S}$-valued random element $X : \Omega \to \mathbb{S}$ if $\mathsf{law}(X_n) \Rightarrow \mathsf{law}(X)$ as $n \to \infty$, which is often denoted by $X_n \Rightarrow X$ or $X_n \xrightarrow{d} X$ as $n \to \infty$.

The following lemma offers a useful equivalent condition for the weak convergence of $\mathbb{D}([0, +\infty))$-valued random elements.

Lemma A.1. *Let $\{X_n \in (\Omega \to \mathbb{D}([0, +\infty))) : n \in \mathbb{N}\}$ and $X : \Omega \to \mathbb{D}([0, +\infty))$ be random elements of $\mathbb{D}([0, +\infty))$. Given any $T \in (0, +\infty)$, let*

$$\pi_T : \mathbb{D}([0, +\infty)) \to \mathbb{D}([0, T]), \quad \pi_T(f) := f|_{[0,T]}.$$

If $\mathbb{P}\{\Delta X(T) := X(T) - X(T-) = 0\} = 1$ for every $T \in (0, +\infty)$, then the following are equivalent:

(i) $X_n \Rightarrow X$ *in* $\mathbb{D}([0,+\infty))$ *as* $n \to \infty$.

(ii) For every $T \in (0,+\infty)$, $\pi_T \circ X_n \Rightarrow \pi_T \circ X$ *in* $\mathbb{D}([0,T])$ *as* $n \to \infty$.

We conclude this section with the following version of *Slutsky's theorem* in the Skorokhod space due to Whitt (1980).

Lemma A.2 (Slutsky's theorem in the Skorokhod space)**.** *If* $X_n \Rightarrow X$ *and* $Y_n \Rightarrow y$ *in the Skorokhod space* $\mathbb{D}(E)$*, where* $y \in \mathbb{C}(E)$ *is deterministic, then we have*

$$X_n + Y_n \Rightarrow X + y \quad \text{and} \quad X_n Y_n \Rightarrow Xy$$

in $\mathbb{D}(E)$*, as* $n \to \infty$.

A.3 Basics of local martingales

Definition A.4 (Adapted stochastic processes)**.** An $\mathbb{S}$-valued stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{S}$ is called $\mathcal{F}_\star$*-adapted* if for every $t \in \mathbb{R}_+$, the $\mathbb{S}$-valued random element $X(t) : \Omega \to \mathbb{S}$ defined by $X(t)(\omega) := X(\omega, t)$ is $\mathcal{F}(t)$-measurable.

Definition A.5 (Predictable stochastic processes)**.**

(i) The $\mathcal{F}_\star$*-predictable* σ*-field* $\mathscr{P}(\mathcal{F}_\star)$ is defined as the smallest σ-field on $\Omega \times \mathbb{R}_+$ making all left-continuous and $\mathcal{F}_\star$-adapted stochastic processes measurable. Equivalently,

$$\mathscr{P}(\mathcal{F}_\star) = \sigma\left(\{A \times (t,+\infty) : t \in \mathbb{R}_+, A \in \mathcal{F}(t)\} \cup \{A \times \{0\} : A \in \mathcal{F}(0)\}\right).$$

(ii) An $\mathbb{S}$-valued stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{S}$ is called $\mathcal{F}_\star$*-predictable* if it is $\mathscr{P}(\mathcal{F}_\star)$-measurable.

Definition A.6 (Martingale)**.** An $\mathcal{F}_\star$-adapted stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{S}$ is an $\mathcal{F}_\star$*-martingale (MG)* if

$$\mathbb{E}[X(t) \mid \mathcal{F}(s)] = X(s)$$

for every $0 \le s \le t < +\infty$.

Definition A.7 (Sub-MG/Super-MG)**.**

(i) A real-valued $\mathcal{F}_\star$-adapted stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is called an $\mathcal{F}_\star$*-sub-martingale* if

$$\mathbb{E}[X(t) \mid \mathcal{F}(s)] \ge X(s)$$

for every $0 \le s \le t < +\infty$.

(ii) A real-valued $\mathcal{F}_\star$-adapted stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is called an $\mathcal{F}_\star$*-super-martingale* if

$$\mathbb{E}[X(t) \mid \mathcal{F}(s)] \le X(s)$$

for every $0 \le s \le t < +\infty$.

Definition A.8 (Stopping time)**.** An $\mathcal{F}_\star$*-stopping time* refers to a random variable $\tau : \Omega \to [0,+\infty]$ such that

$$\tau^{-1}([0,t]) := \{\omega \in \Omega : \tau(\omega) \in [0,t]\} \in \mathcal{F}(t)$$

for every $t \in \mathbb{R}_+$.

Definition A.9 (Local martingale)**.** An $\mathcal{F}_\star$-adapted stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{S}$ is called an $\mathcal{F}_\star$*-local martingale* if there exists a sequence $\{\tau_k \in (\Omega \to \mathbb{R}_+) : k \in \mathbb{Z}_+\}$ such that

(i) the τ_k's are $\mathbb{P}$-almost surely increasing: $\mathbb{P}\{\tau_k < \tau_{k+1}\} = 1$ for every $k \in \mathbb{Z}_+$.

(ii) $\{\tau_k : k \in \mathbb{Z}_+\}$ diverges $\mathbb{P}$-almost surely: $\mathbb{P}\{\lim_{k\to\infty} \tau_k = +\infty\} = 1$.

(iii) the τ_k-*stopped process* $X^{\tau_k} : \Omega \times \mathbb{R}_+ \to \mathbb{S}$, where

$$X^{\tau_k}(\omega, t) := X\left(\omega, \min\left\{t, \tau_k(\omega)\right\}\right), \quad (k, \omega, t) \in \mathbb{Z}_+ \times \Omega \times [0, +\infty),$$

is an $\mathcal{F}_\star$-martingale for every $k \in \mathbb{Z}_+$.

Definition A.10 (Local sub-MG/local super-MG)**.**

(i) A real-valued $\mathcal{F}_\star$-adapted stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is called an $\mathcal{F}_\star$-*local sub-martingale* if there exists a sequence $\{\tau_k \in (\Omega \to \mathbb{R}_+) : k \in \mathbb{Z}_+\}$ such that

- the τ_k's are $\mathbb{P}$-almost surely increasing: $\mathbb{P}\{\tau_k < \tau_{k+1}\} = 1$ for every $k \in \mathbb{Z}_+$.
- $\{\tau_k : k \in \mathbb{Z}_+\}$ diverges $\mathbb{P}$-almost surely: $\mathbb{P}\{\lim_{k\to\infty} \tau_k = +\infty\} = 1$.
- the τ_k-stopped process $X^{\tau_k} : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is an $\mathcal{F}_\star$-sub-martingale for every $k \in \mathbb{Z}_+$.

(ii) A real-valued $\mathcal{F}_\star$-adapted stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is called an $\mathcal{F}_\star$-*local super-martingale* if there exists a sequence $\{\tau_k \in (\Omega \to \mathbb{R}_+) : k \in \mathbb{Z}_+\}$ such that

- the τ_k's are $\mathbb{P}$-almost surely increasing: $\mathbb{P}\{\tau_k < \tau_{k+1}\} = 1$ for every $k \in \mathbb{Z}_+$.
- $\{\tau_k : k \in \mathbb{Z}_+\}$ diverges $\mathbb{P}$-almost surely: $\mathbb{P}\{\lim_{k\to\infty} \tau_k = +\infty\} = 1$.
- the τ_k-stopped process $X^{\tau_k} : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is an $\mathcal{F}_\star$-super-martingale for every $k \in \mathbb{Z}_+$.

We now consider any square-integrable $\mathcal{F}_\star$-local martingale $M : \Omega \times \mathbb{R}_+ \to \mathbb{R}$. As $M^2 : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$ is an $\mathcal{F}_\star$-local sub-martingale, the Doob-Meyer decomposition theorem from Dellacherie & Meyer (1982) guarantees the existence of a unique non-decreasing $\mathcal{F}_\star$-predictable stochastic process $\langle M \rangle : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ with $\langle M \rangle(0) = 0$ and $M^2 - \langle M \rangle : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is an $\mathcal{F}_\star$-local-martingale. The unique non-decreasing $\mathcal{F}_\star$-predictable stochastic process $\langle M \rangle : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is often called the $\mathcal{F}_\star$-*predictable quadratic variation* of the square-integrable $\mathcal{F}_\star$-local martingale $M : \Omega \times \mathbb{R}_+ \to \mathbb{R}$.

Definition A.11 (Càdlàg stochastic process)**.** Let $(\mathcal{M}, \rho)$ be a metric space. An $\mathcal{M}$-valued stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathcal{M}$ is said to be *càdlàg* if $X(\omega, \cdot) \in \mathbb{D}([0, +\infty), \mathcal{M})$ for every $\omega \in \Omega$.

Definition A.12 (Jump)**.** Let $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ be any real-valued stochastic process. The *jump* of $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ refers to a real-valued stochastic process $\Delta X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ defined as

$$\Delta X(0) = 0 \quad \text{and} \quad \Delta X(t) = X(t) - X(t-), \quad t \in (0, +\infty).$$

Definition A.13 (Pure-jump process)**.** A *pure-jump stochastic process* refers to a real-valued càdlàg stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ that can be written as

$$X(t) = X(0) + \sum_{s \in (0,t]} \Delta X(s), \quad t \in \mathbb{R}_+, \tag{A.3}$$

where the summation over $s \in (0, t]$ in equation A.3 is valid owing to the fact that the sample trajectory $X(\omega, \cdot) : \mathbb{R}_+ \to \mathbb{R}$ is a càdlàg function implies the at most countability of the set of discontinuity points

$$\mathcal{D}\{X(\omega, \cdot)\} := \{t \in \mathbb{R}_+ : \Delta X(\omega, t) > 0\}$$

for every $\omega \in \Omega$.

Lemma A.3 (Predictable quadratic variation of pure-jump local martingales)**.** *Let $M : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ be a pure-jump, square-integrable $\mathcal{F}_\star$-local martingale. Then, its $\mathcal{F}_\star$-predictable quadratic variation $\langle M \rangle : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is given by*

$$\langle M \rangle(t) = \sum_{s \in (0,t]} \mathbb{E}\left[\{\Delta M(s)\}^2 \middle| \mathcal{F}(s-)\right] \quad \mathbb{P}\text{-almost surely} \tag{A.4}$$

for every $t \in \mathbb{R}_+$.

A.4 A FEW PRELIMINARY FACTS

In this section, we collect a couple of useful preliminary facts from the literature of probability theory and stochastic processes that facilitates our analysis.

Lemma A.4 (Cantelli's inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space and $X : \Omega \to \mathbb{R}$ be a square-integrable random variable, i.e., $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Then, it holds for any $\lambda \in \mathbb{R}_+$ that*

$$\max\left\{\mathbb{P}\left\{X - \mathbb{E}[X] \geq \lambda\right\}, \mathbb{P}\left\{X - \mathbb{E}[X] \leq -\lambda\right\}\right\} \leq \frac{\sigma^2}{\sigma^2 + \lambda^2} \tag{A.5}$$

where $\sigma^2 := \mathrm{Var}[X] \in \mathbb{R}_+$ denotes the variance of X.

Lemma A.5 (Multivariate Cochran's theorem Wong et al. (1991))**.** *Let $\mathbb{X}_n \in \mathbb{R}^{n \times d}$ be a random matrix with i.i.d. row vectors*

$$\{\mathbf{X}_i : i \in [n]\} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(\mathbf{0}_d, \mathbf{I}_d).$$

Let $\mathbf{Q}_1, \mathbf{Q}_2, \cdots, \mathbf{Q}_k \in \mathcal{S}^n$ be symmetric idempotent $n \times n$ matrices (i.e., $\mathbf{Q}_i^2 = \mathbf{Q}_i$ for every $i \in [k]$) such that

$$\sum_{i=1}^{k} \mathbf{Q}_i = \mathbf{I}_n \quad \text{and} \quad \sum_{i=1}^{k} \mathrm{rank}(\mathbf{Q}_i) = n. \tag{A.6}$$

Then, it follows that $\mathbf{S}_i := \mathbb{X}_n^\top \mathbf{Q}_i \mathbb{X}_n \sim \mathbb{W}_d(\mathrm{rank}(\mathbf{Q}_i), \mathbf{I}_d)$ for every $i \in [k]$, and $\{\mathbf{S}_i : i \in [k]\}$ are independent $d \times d$ random matrices.

Lemma A.6 (*Theorem 4.2.12* in Durrett (2019))**.** *Let $\mathcal{F}_\star := (\mathcal{F}_n : n \in \mathbb{Z}_+)$ be a filtration and $(X_n : n \in \mathbb{Z}_+)$ be a non-negative $\mathcal{F}_\star$-super-martingale defined on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, there exists a random variable $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \overset{n \to \infty}{\longrightarrow} X_\infty$ $\mathbb{P}$-almost surely and $0 \leq \mathbb{E}[X_\infty] \leq \mathbb{E}[X_0]$.*

Lemma A.7 (Doob's L^p-martingale convergence theorem: *Theorem 4.4.6* in Durrett (2019))**.** *Let $\mathcal{F}_\star := (\mathcal{F}_n : n \in \mathbb{Z}_+)$ be a filtration and $(X_n : n \in \mathbb{Z}_+)$ be an L^p-$\mathcal{F}_\star$-martingale defined on $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., an $\mathcal{F}_\star$-martingale such that $\sup\{\mathbb{E}[|X_n|^p] : n \in \mathbb{Z}_+\} < +\infty$, where $p \in (1, +\infty)$. Then, there is a random variable $X_\infty \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \overset{n \to \infty}{\longrightarrow} X_\infty$ $\mathbb{P}$-almost surely and in L^p.*

The ordinary continuous mapping theorem argues that a *fixed* continuous map preserves modes of convergence of a sequence of random elements in metric spaces. The following result due to van der Vaart & Wellner (1996) says that the same conclusion remains true even when the transformation is allowed to change with $n \in \mathbb{N}$, write f_n, provided that the sequence $\{f_n : n \in \mathbb{N}\}$ behaves asymptotically like a limit map f along every convergent deterministic sequence $x_n \to x$ near the support of the limit of a sequence of random elements.

Lemma A.8 (The extended continuous mapping theorem)**.** *Let $(\mathbb{D}, d)$ and $(\mathbb{E}, e)$ be two metric spaces. For each $n \in \mathbb{N}$, let $\mathbb{D}_n \in \mathcal{B}(\mathbb{D})$ and $f_n : \mathbb{D}_n \to \mathbb{E}$, and let $\mathbb{D}_0 \in \mathcal{B}(\mathbb{D})$ and $f : \mathbb{D}_0 \to \mathbb{E}$. Also, we consider a sequence of $\mathbb{D}$-valued random elements $\{X_n \in (\Omega \to \mathbb{D}_n) : n \in \mathbb{N}\}$ and $X : \Omega \to \mathbb{D}$. Suppose that*

- *(i) $X_n \Rightarrow X$ as $n \to \infty$ in $\mathbb{D}$.*
- *(ii) $\mathbb{D}_0$ is a separable subset of $\mathbb{D}$ under the metric topology induced by d, and $\mathbb{P}\{X \in \mathbb{D}_0\} = 1$.*
- *(iii) whenever $x_n \in \mathbb{D}_n$ and $x \in \mathbb{D}_0$ satisfy $x_n \to x$ as $n \to \infty$ in $\mathbb{D}$, one has $f_n(x_n) \to f(x)$ as $n \to \infty$ in $\mathbb{E}$.*

Then, it holds that $f_n(X_n) \Rightarrow f(X)$ as $n \to \infty$ in $\mathbb{E}$.

The next result is a martingale functional central limit theorem due to Rebolledo (1980), which can be viewed as a *functional invariance principle* for local martingales. When a sequence of local martingales accumulates conditional variance in a deterministic and well-behaved manner, and none of its individual jumps is large enough to affect the limit, then its entire sample path converges

weakly to a *scaled* standard one-dimensional Brownian motion. It therefore turns the problem of establishing a functional CLT into two verifiable tasks: (i) identify the limits of predictable quadratic variations, and (ii) verify a *Lindeberg condition* holds, i.e., the local martingale has no macroscopic jumps.

Lemma A.9 (The continuous-time martingale functional central limit theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the underlying probability space. For each $n \in \mathbb{N}$, let $\mathcal{F}_n(\star) := (\mathcal{F}_n(t) : t \in \mathbb{R}_+)$ be a filtration defined on $(\Omega, \mathcal{F})$ and let $M_n : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ be a càdlàg, square-integrable local $\mathcal{F}_n(\star)$-martingale with $M_n(0) = 0$. Let $\langle M_n \rangle : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ denote the $\mathcal{F}_n(\star)$-predictable quadratic variation and $\Delta M_n : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ denote the jump of $M_n : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ for each $n \in \mathbb{N}$. Now, suppose for some $T \in (0, +\infty)$ that:*

(i) **Predictable quadratic variation convergence***: There exists some universal constant $c \in (0, +\infty)$ such that*

$$\sup \{|\langle M_n \rangle (t) - ct| : t \in [0, T]\} \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{A.7}$$

(ii) **Lindeberg condition***: For every $\varepsilon \in (0, +\infty)$, one has*

$$\sum_{t \in (0,T]} \mathbb{E}\left[\{\Delta M_n(t)\}^2 \cdot \mathbb{1}_{\{|\Delta M_n(t)| > \varepsilon\}} \middle| \mathcal{F}_n(t-) \right] \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{A.8}$$

Then, it holds that $\pi_T \circ M_n \Rightarrow c\,(\pi_T \circ \mathbb{B})$ as $n \to \infty$ in $\mathbb{D}([0, T])$.

B Omitted details and proofs for Section 2

B.1 Proof of Proposition 2.1

It holds that

$$\begin{aligned}
&\mathcal{R}(f_{\mathbf{w}}) \\
&= \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \{Y \neq f_{\mathbf{w}}(\mathbf{X})\} \\
&= \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \{f_{\mathbf{w}}(\mathbf{X}) = -1 \mid Y = +1\} \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \{Y = +1\} \\
&\quad + \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \{f_{\mathbf{w}}(\mathbf{X}) = +1 \mid Y = -1\} \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \{Y = -1\} \\
&= \frac{1}{2} \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left\{ \mathbf{w}^\top \mathbf{X} < 0 \mid Y = +1 \right\} + \frac{1}{2} \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left\{ \mathbf{w}^\top \mathbf{X} \geq 0 \mid Y = -1 \right\} \\
&= \frac{1}{2} \mathbb{P}_{\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left\{ \mathbf{w}^\top \mathbf{X} < 0 \right\} + \frac{1}{2} \mathbb{P}_{\mathbf{X} \sim \mathcal{N}_d(-\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left\{ \mathbf{w}^\top \mathbf{X} \geq 0 \right\} \\
&\overset{\text{(a)}}{=} \frac{1}{2} \mathbb{P}_{Z \sim \mathcal{N}(\mathbf{w}^\top \boldsymbol{\mu}_\star, \mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w})} \{Z < 0\} + \frac{1}{2} \mathbb{P}_{Z \sim \mathcal{N}(-\mathbf{w}^\top \boldsymbol{\mu}_\star, \mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w})} \{Z \geq 0\} \\
&= \mathbb{P}_{Z \sim \mathcal{N}(\mathbf{w}^\top \boldsymbol{\mu}_\star, \mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w})} \{Z \leq 0\} \\
&= \mathbb{P}_{Z \sim \mathcal{N}(\mathbf{w}^\top \boldsymbol{\mu}_\star, \mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w})} \left\{ \frac{Z - \mathbf{w}^\top \boldsymbol{\mu}_\star}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w}}} \leq -\frac{\mathbf{w}^\top \boldsymbol{\mu}_\star}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w}}} \right\} \\
&= \Phi\left(-\frac{\mathbf{w}^\top \boldsymbol{\mu}_\star}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w}}} \right),
\end{aligned}$$

where the step (a) holds since

$$\mathbf{X} \sim \mathcal{N}_d(y \boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \quad \Rightarrow \quad \mathbf{w}^\top \mathbf{X} \sim \mathcal{N}\left(y \mathbf{w}^\top \boldsymbol{\mu}_\star, \mathbf{w}^\top \boldsymbol{\Sigma}_\star \mathbf{w}\right)$$

for every $y \in \{\pm 1\}$, concluding the proof of Proposition 2.1.

B.2 Iterative model training via maximum likelihood estimation (MLE)

In this section, we present the precise procedure for iterative model fitting via MLE. Specifically, we provide a comprehensive description for the discard-workflow for iterative MLE.

Algorithm 2 Iterative MLE: the discard-workflow

Input: A sequence of numbers of synthetic samples $(n_t \in \mathbb{N} : t \in \mathbb{Z}_+)$ and the number of training generations $T \in \mathbb{N}$.

1: Set $\hat{\boldsymbol{\theta}}_0 := \boldsymbol{\theta}_\star \in \Theta$.
2: **for** $t = 0, 1, 2, \cdots, T-1$ **do**:
3: Sample $\mathbf{X}_{1:n_t}^{(t)} := \left(X_i^{(t)} : i \in [n_t]\right) \sim \mathbb{P}_{\hat{\boldsymbol{\theta}}_t}^{\otimes n_t}$.
4: Define the negative log-likelihood function $\mathcal{L}_{n_t}^{(t)} : \Theta \to \mathbb{R}$ at the t-th generation by

$$\mathcal{L}_{n_t}^{(t)}(\boldsymbol{\theta}) := -\frac{1}{n_t} \sum_{i=1}^{n_t} \log p_{\boldsymbol{\theta}} \left(X_i^{(t)}\right). \tag{B.1}$$

5: Train model on the synthetically generated data $\mathbf{X}_{1:n_t}^{(t)}$ via MLE:

$$\hat{\boldsymbol{\theta}}_{t+1} \in \arg\min \left\{ \mathcal{L}_{n_t}^{(t)}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \right\}. \tag{B.2}$$

6: **end for**
7: **return** the model parameter $\hat{\boldsymbol{\theta}}_T \in \Theta$.

C Omitted details and proofs for Section 3

C.1 Proof of Proposition 3.1

To begin with, let us recall the evolution of the model parameters under the discard-workflow of iterative MLE under the symmetric binary GMM equation 2.1 with the constant size of synthetic samples $n_t \equiv n$: for any fixed generation step $t \in \mathbb{Z}_+$,

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_{t+1}(n) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^{(t)} Y_i^{(t)}, \\ \widehat{\boldsymbol{\Sigma}}_{t+1}(n) &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbf{X}_i^{(t)} - Y_i^{(t)} \widehat{\boldsymbol{\mu}}_{t+1}(n) \right\} \left\{ \mathbf{X}_i^{(t)} - Y_i^{(t)} \widehat{\boldsymbol{\mu}}_{t+1}(n) \right\}^\top. \end{aligned} \tag{C.1}$$

Note that if $\mathbf{O}_{1:n}^{(t)} = \left(O_i^{(t)} = \left(\mathbf{X}_i^{(t)}, Y_i^{(t)}\right) : i \in [n]\right) \sim \mathbb{P}_{\left(\widehat{\boldsymbol{\mu}}_t(n), \widehat{\boldsymbol{\Sigma}}_t(n)\right)}^{\otimes n}$, then we have

$$\begin{aligned} \mathbf{Y}_{1:n}^{(t)} &= \left(Y_i^{(t)} : i \in [n]\right) \sim \text{Unif}\left(\{\pm 1\}^n\right), \\ \mathbf{X}_i^{(t)} &= Y_i^{(t)} \widehat{\boldsymbol{\mu}}_t(n) + \widehat{\boldsymbol{\Sigma}}_t^{\frac{1}{2}}(n) \mathbf{Z}_i^{(t)}, \quad \forall i \in [n], \end{aligned} \tag{C.2}$$

where $\left(\mathbf{Z}_i^{(t)} : i \in [n]\right) \sim \left\{\mathcal{N}_d\left(\mathbf{0}_d, \mathbf{I}_d\right)\right\}^{\otimes n}$. Taking two equations equation C.1 and equation C.2 collectively yields

$$\widehat{\boldsymbol{\mu}}_{t+1}(n) = \frac{1}{n} \sum_{i=1}^{n} Y_i^{(t)} \left[Y_i^{(t)} \widehat{\boldsymbol{\mu}}_t(n) + \widehat{\boldsymbol{\Sigma}}_t^{\frac{1}{2}}(n) \mathbf{Z}_i^{(t)} \right] = \widehat{\boldsymbol{\mu}}_t(n) + \frac{1}{n} \widehat{\boldsymbol{\Sigma}}_t^{\frac{1}{2}}(n) \sum_{i=1}^{n} Y_i^{(t)} \mathbf{Z}_i^{(t)}, \tag{C.3}$$

and

$$\begin{aligned} &\widehat{\boldsymbol{\Sigma}}_{t+1}(n) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbf{X}_i^{(t)} - Y_i^{(t)} \widehat{\boldsymbol{\mu}}_{t+1}(n) \right\} \left\{ \mathbf{X}_i^{(t)} - Y_i^{(t)} \widehat{\boldsymbol{\mu}}_{t+1}(n) \right\}^\top \\ &= \widehat{\boldsymbol{\Sigma}}_t^{\frac{1}{2}}(n) \underbrace{\left[\frac{1}{n} \sum_{j=1}^{n} \left\{ \mathbf{Z}_j^{(t)} - \frac{1}{n} Y_j^{(t)} \sum_{i=1}^{n} Y_i^{(t)} \mathbf{Z}_i^{(t)} \right\} \left\{ \mathbf{Z}_j^{(t)} - \frac{1}{n} Y_j^{(t)} \sum_{i=1}^{n} Y_i^{(t)} \mathbf{Z}_i^{(t)} \right\}^\top \right]}_{=: \mathbf{S}_t(n)} \widehat{\boldsymbol{\Sigma}}_t^{\frac{1}{2}}(n). \end{aligned} \tag{C.4}$$

It is easy to observe that if $((Y_i, \mathbf{Z}_i) : i \in [n]) \sim \{\text{Unif}(\{\pm 1\}) \otimes \mathcal{N}_d(\mathbf{0}_d, \mathbf{I}_d)\}^{\otimes n}$, then

$$(Y_i \mathbf{Z}_i : i \in [n]) \sim \{\mathcal{N}_d(\mathbf{0}_d, \mathbf{I}_d)\}^{\otimes n} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} Y_i \mathbf{Z}_i \sim \mathcal{N}_d\left(\mathbf{0}_d, \frac{1}{n}\mathbf{I}_d\right).$$

Thus, the equation C.3 gives

$$\widehat{\boldsymbol{\mu}}_{T+1}(n) | \, \mathcal{V}_T(n) \sim \mathcal{N}_d\left(\widehat{\boldsymbol{\mu}}_T(n), \frac{1}{n}\widehat{\boldsymbol{\Sigma}}_T(n)\right), \tag{C.5}$$

concluding the proof of the part (i) of Proposition 3.1.

We now turn to establish the part (ii) of Proposition 3.1. Towards this end, we take a closer inspection on the $d \times d$ random matrix $\mathbf{S}_t(n)$ defined as

$$\mathbf{S}_t(n) := \frac{1}{n}\sum_{j=1}^{n}\left\{\mathbf{Z}_j^{(t)} - \frac{1}{n}Y_j^{(t)}\sum_{i=1}^{n} Y_i^{(t)}\mathbf{Z}_i^{(t)}\right\}\left\{\mathbf{Z}_j^{(t)} - \frac{1}{n}Y_j^{(t)}\sum_{i=1}^{n} Y_i^{(t)}\mathbf{Z}_i^{(t)}\right\}^{\top}. \tag{C.6}$$

For notational simplicity, let $\mathbb{Z}_n^{(t)} := \left[\begin{array}{c|c|c|c} \mathbf{Z}_1^{(t)} & \mathbf{Z}_2^{(t)} & \cdots & \mathbf{Z}_n^{(t)} \end{array}\right]^{\top} \in \mathbb{R}^{n \times d}$. Then, one can reveal that

$$n \cdot \mathbf{S}_t(n) := \left\{\mathbb{Z}_n^{(t)}\right\}^{\top} \underbrace{\left[\sum_{j=1}^{n}\left\{\mathbf{e}_j^n - \frac{1}{n}Y_j^{(t)}\mathbf{Y}_{1:n}^{(t)}\right\}\left\{\mathbf{e}_j^n - \frac{1}{n}Y_j^{(t)}\mathbf{Y}_{1:n}^{(t)}\right\}^{\top}\right]}_{=: \, \mathbf{M}_n\left(\mathbf{Y}_{1:n}^{(t)}\right)} \mathbb{Z}_n^{(t)}, \tag{C.7}$$

where $\mathbf{e}_j^n \in \mathbb{R}^n$ denotes the vector with a 1 in the j-th coordinate and 0's elsewhere for each $j \in [n]$, and the function $\mathbf{M}_n : \{\pm 1\}^n \to \mathbb{S}_+^n$ is defined as

$$\begin{aligned} \mathbf{M}_n(\mathbf{y}_{1:n}) &:= \sum_{j=1}^{n}\left\{\mathbf{e}_j^n - \frac{1}{n}y_j\mathbf{y}_{1:n}\right\}\left\{\mathbf{e}_j^n - \frac{1}{n}y_j\mathbf{y}_{1:n}\right\}^{\top} \\ &= \underbrace{\sum_{j=1}^{n}\mathbf{e}_j^n\left(\mathbf{e}_j^n\right)^{\top}}_{=\mathbf{I}_n} - \frac{1}{n}\underbrace{\left(\sum_{j=1}^{n} y_j\mathbf{e}_j^n\right)}_{=\mathbf{y}_{1:n}}\mathbf{y}_{1:n}^{\top} \\ &\quad - \frac{1}{n}\mathbf{y}_{1:n}\left(\sum_{j=1}^{n} y_j\mathbf{e}_j^n\right)^{\top} + \frac{1}{n}\mathbf{y}_{1:n}\mathbf{y}_{1:n}^{\top} \\ &= \mathbf{I}_n - \frac{1}{n}\mathbf{y}_{1:n}\mathbf{y}_{1:n}^{\top}. \end{aligned} \tag{C.8}$$

From the equation equation C.8, it follows that $\text{rank}\{\mathbf{M}_n(\mathbf{y}_{1:n})\} = n-1$ and $\{\mathbf{M}_n(\mathbf{y}_{1:n})\}^2 = \mathbf{M}_n(\mathbf{y}_{1:n})$ for every $\mathbf{y}_{1:n} \in \{\pm 1\}^n$. Thus, the multivariate Cochran's theorem (Lemma A.5) implies $n \cdot \mathbf{S}_t(n) \sim \mathbb{W}_d(n-1, \mathbf{I}_d)$. Hence, taking this fact together with the representation equation C.4 collectively completes the proof of the part (ii) of Proposition 3.1. Note that the equation equation 3.3 is an immediate consequence of the part (ii) of Proposition 3.1.

Finally, it only remains to deduce the fact equation 3.2. If we define $\boldsymbol{\Delta}_t(n) := \widehat{\boldsymbol{\mu}}_{t+1}(n) - \widehat{\boldsymbol{\mu}}_t(n)$ for $t \in \mathbb{Z}_+$, then the part (i) of Proposition 3.1 implies

$$\boldsymbol{\Delta}_t(n) | \, \mathcal{F}_t(n) \sim \mathcal{N}_d\left(\mathbf{0}_d, \frac{1}{n}\widehat{\boldsymbol{\Sigma}}_t(n)\right), \quad t \in \mathbb{Z}_+. \tag{C.9}$$

Hence, the desired claim equation 3.2 directly follows as a consequence of $\widehat{\boldsymbol{\mu}}_T(n) = \boldsymbol{\mu}_\star + \sum_{t=0}^{T-1}\boldsymbol{\Delta}_t(n)$ for each $T \in \mathbb{Z}_+$ and this completes the proof of Proposition 3.1.

C.2 Proof of Proposition 3.2

It suffices to show that $\left\|\widehat{\mathbf{\Sigma}}_T(n)\right\|_{\text{op}} \overset{T\to\infty}{\longrightarrow} 0$ $\mathbb{P}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}$-almost surely to establish the $\mathbb{P}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}$-almost sure convergence of $\widehat{\mathbf{\Sigma}}_T(n)$ to $\mathbf{O}_{d\times d}$ as $T\to\infty$. Define $M_t(n) := \left(1-\frac{1}{n}\right)^{-t} \text{Trace}\left\{\widehat{\mathbf{\Sigma}}_t(n)\right\}$ for $t\in\mathbb{Z}_+$. As it turns out, $(M_t(n) : t\in\mathbb{Z}_+)$ is a non-negative $\mathcal{F}_\star(n)$-martingale for any fixed sample size $n>1$, since

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[M_{t+1}(n)\mid\mathcal{F}_t(n)\right] &= \left(1-\frac{1}{n}\right)^{-(t+1)} \text{Trace}\left(\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\widehat{\mathbf{\Sigma}}_{t+1}(n)\middle|\mathcal{F}_t(n)\right]\right)\\ &\overset{\text{(a)}}{=} \left(1-\frac{1}{n}\right)^{-t} \text{Trace}\left\{\widehat{\mathbf{\Sigma}}_t(n)\right\}\\ &= M_t(n),\end{aligned} \tag{C.10}$$

where the step (a) is a result of the conditional contraction identity equation 3.3. Thus, Lemma A.6 implies that there is an integrable random variable $M_\infty(n)\in[0,+\infty)$ with $M_T(n)\to M_\infty(n)$ $\mathbb{P}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}$-almost surely as $T\to\infty$. Hence, it follows by letting $T\to\infty$ that

$$\text{Trace}\left\{\widehat{\mathbf{\Sigma}}_T(n)\right\} = \left(1-\frac{1}{n}\right)^{T} M_T(n) \to 0\cdot M_\infty(n) = 0 \quad \mathbb{P}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\text{-almost surely.} \tag{C.11}$$

This immediately yields the desired conclusion $\left\|\widehat{\mathbf{\Sigma}}_T(n)\right\|_{\text{op}}\to 0$ $\mathbb{P}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}$-almost surely as $T\to\infty$, since the fact that $\widehat{\mathbf{\Sigma}}_t(n)\in\mathcal{S}_+^d$ for every $t\in\mathbb{Z}_+$ implies

$$\left\|\widehat{\mathbf{\Sigma}}_T(n)\right\|_{\text{op}} \le \text{Trace}\left\{\widehat{\mathbf{\Sigma}}_T(n)\right\} \to 0 \quad \mathbb{P}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\text{-almost surely}$$

as $T\to\infty$, which finishes the proof of Proposition 3.2.

C.3 Proof of Theorem 3.1

We embark on the proof of Theorem 3.1 with the following key observations directly obtained from the part (i) of Proposition 3.1:

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_{t+1}(n)\middle|\,\mathcal{F}_t(n)\right] &= \widehat{\boldsymbol{\mu}}_t(n) \quad \text{and}\\ \text{Cov}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_{t+1}(n)\middle|\,\mathcal{F}_t(n)\right] &= \frac{1}{n}\widehat{\mathbf{\Sigma}}_t(n),\end{aligned} \tag{C.12}$$

and in particular, $(\widehat{\boldsymbol{\mu}}_t(n) : t\in\mathbb{Z}_+)$ forms a $\mathcal{F}_\star(n)$-martingale. We now claim that $(\widehat{\boldsymbol{\mu}}_t(n) : t\in\mathbb{Z}_+)$ forms an L^2-$\mathcal{F}_\star(n)$-martingale, i.e., an $\mathcal{F}_\star(n)$-martingale such that

$$\sup\left\{\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\|\widehat{\boldsymbol{\mu}}_t(n)\|_2^2\right] : t\in\mathbb{Z}_+\right\} < +\infty. \tag{C.13}$$

As $(\boldsymbol{\Delta}_t(n) : t\in\mathbb{Z}_+)$ forms a $\mathcal{F}_\star(n)$-martingale difference sequence, it follows for any $0\le s<t<+\infty$ that

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\{\boldsymbol{\Delta}_s(n)\}^\top\boldsymbol{\Delta}_t(n)\right] \overset{\text{(a)}}{=} \mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\{\boldsymbol{\Delta}_s(n)\}^\top\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\boldsymbol{\Delta}_t(n)\middle|\,\mathcal{F}_t(n)\right]\right] = 0, \tag{C.14}$$

where the step (a) follows by the $\mathcal{F}_t(n)$-measurability of $\boldsymbol{\Delta}_s(n)$. Writing $\widehat{\boldsymbol{\mu}}_t(n) = \boldsymbol{\mu}_\star + \sum_{s=0}^{t-1}\boldsymbol{\Delta}_s(n)$ for each $t\in\mathbb{Z}_+$, the L^2-orthogonality property equation C.14 of the increments $(\boldsymbol{\Delta}_t(n) : t\in\mathbb{Z}_+)$ yields

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\|\widehat{\boldsymbol{\mu}}_t(n)\|_2^2\right] &= \|\boldsymbol{\mu}_\star\|_2^2 + \sum_{s=0}^{t-1}\mathbb{E}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\|\boldsymbol{\Delta}_s(n)\|_2^2\right]\\ &= \|\boldsymbol{\mu}_\star\|_2^2 + \sum_{s=0}^{t-1}\text{Trace}\left\{\text{Cov}_{(\boldsymbol{\mu}_\star,\mathbf{\Sigma}_\star)}\left[\boldsymbol{\Delta}_s(n)\right]\right\}\end{aligned}$$

$$
\begin{aligned}
&\overset{\text{(b)}}{=} \|\boldsymbol{\mu}_\star\|_2^2 + \sum_{s=0}^{t-1} \text{Trace}\left\{\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\boldsymbol{\Delta}_s(n)\middle|\mathcal{F}_s(n)\right]\right]\right\} \\
&\quad + \sum_{s=0}^{t-1} \text{Trace}\left\{\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\boldsymbol{\Delta}_s(n)\middle|\mathcal{F}_s(n)\right]\right]\right\} \qquad \text{(C.15)} \\
&\overset{\text{(c)}}{=} \|\boldsymbol{\mu}_\star\|_2^2 + \frac{1}{n}\sum_{s=0}^{t-1} \text{Trace}\left\{\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\hat{\boldsymbol{\Sigma}}_s(n)\right]\right\} \\
&\overset{\text{(d)}}{=} \|\boldsymbol{\mu}_\star\|_2^2 + \frac{1}{n}\sum_{s=0}^{t-1} \left(1-\frac{1}{n}\right)^s \text{Trace}\left(\boldsymbol{\Sigma}_\star\right) \\
&= \|\boldsymbol{\mu}_\star\|_2^2 + \left\{1-\left(1-\frac{1}{n}\right)^t\right\} \text{Trace}\left(\boldsymbol{\Sigma}_\star\right),
\end{aligned}
$$

where the step (b) follows due to the law of total expectation, the step (c) invokes the facts equation C.12, and the step (d) results from the identity equation 3.3. Hence, it follows from the observation equation C.15 that

$$
\begin{aligned}
&\sup\left\{\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\|\widehat{\boldsymbol{\mu}}_t(n)\|_2^2\right] : t \in \mathbb{Z}_+\right\} \\
&= \lim_{t\to\infty}\left[\|\boldsymbol{\mu}_\star\|_2^2 + \left\{1-\left(1-\frac{1}{n}\right)^t\right\} \text{Trace}\left(\boldsymbol{\Sigma}_\star\right)\right] \qquad \text{(C.16)} \\
&= \|\boldsymbol{\mu}_\star\|_2^2 + \text{Trace}\left(\boldsymbol{\Sigma}_\star\right) < +\infty,
\end{aligned}
$$

as desired. Applying Doob's L^2-martingale convergence theorem (Lemma A.7 with $p = 2$) guarantees the existence of a d-dimensional random vector $\widehat{\boldsymbol{\mu}}_\infty(n) \in L^2\left(\mathbb{R}^d \times \{\pm 1\}, \mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\right)$ such that as $T \to \infty$, $\widehat{\boldsymbol{\mu}}_T(n) \to \widehat{\boldsymbol{\mu}}_\infty(n)$ $\mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}$-almost surely and in L^2.

Finally, let us turn our attention to characterize the moments of $\widehat{\boldsymbol{\mu}}_\infty(n) \in L^2\left(\mathbb{R}^d \times \{\pm 1\}, \mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\right)$ up to the second-order. Since the L^2-convergence automatically implies the L^1-convergence, we have

$$
\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_\infty(n)\right] = \lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_T(n)\right] \overset{\text{(e)}}{=} \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_0(n)\right] = \boldsymbol{\mu}_\star,
$$

where the step (e) holds since $(\widehat{\boldsymbol{\mu}}_t(n) : t \in \mathbb{Z}_+)$ is a $\mathcal{F}_\star(n)$-martingale. On the other hand, one has

$$
\begin{aligned}
&\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_{t+1}(n)\right] \\
&\overset{\text{(f)}}{=} \text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_{t+1}(n)\middle|\mathcal{F}_t(n)\right]\right] + \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_{t+1}(n)\middle|\mathcal{F}_t(n)\right]\right] \qquad \text{(C.17)} \\
&\overset{\text{(g)}}{=} \text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_t(n)\right] + \frac{1}{n}\left(1-\frac{1}{n}\right)^t \boldsymbol{\Sigma}_\star,
\end{aligned}
$$

where the step (f) utilizes the law of total expectation, and the step (g) follows due to the facts equation C.12. The identity equation C.17 then yields

$$
\begin{aligned}
&\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\widehat{\boldsymbol{\mu}}_T(n)\right] \\
&= \sum_{t=0}^{T-1}\left\{\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\hat{\boldsymbol{\mu}}_{t+1}(n)\right] - \text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\hat{\boldsymbol{\mu}}_t(n)\right]\right\} + \underbrace{\text{Cov}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\hat{\boldsymbol{\mu}}_0(n)\right]}_{=\mathbf{0}_d} \\
&= \sum_{t=0}^{T-1} \frac{1}{n}\left(1-\frac{1}{n}\right)^t \boldsymbol{\Sigma}_\star \qquad \text{(C.18)} \\
&= \left\{1-\left(1-\frac{1}{n}\right)^T\right\} \boldsymbol{\Sigma}_\star.
\end{aligned}
$$

Hence, the L^2-convergence $\widehat{\boldsymbol{\mu}}_T(n) \xrightarrow{L^2} \widehat{\boldsymbol{\mu}}_\infty(n)$ as $T \to \infty$ together with the finding equation C.18 implies

$$\mathrm{Cov}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\hat{\boldsymbol{\mu}}_\infty(n)\right] = \lim_{T\to\infty} \mathrm{Cov}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\hat{\boldsymbol{\mu}}_T(n)\right] = \boldsymbol{\Sigma}_\star,$$

which completes the proof of Theorem 3.1.

D Omitted details and proofs for Section 4

D.1 Several key lemmas

Before proceeding, let us introduce the following result that characterizes the (unconditional) law of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$, whose proof is deferred to Appendix D.5.

Lemma D.1 (Law of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$)**.** *Given any initialization of the ground-truth model parameter $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$, the training dynamics generated by the discard-workflow of iterative MLE under the symmetric binary GMM equation 2.1 (Algorithm 1) satisfies the following property in terms of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$: for any fixed sample size $n \geq d+1$, it holds that*

$$\widehat{\boldsymbol{\mu}}_\infty(n) \overset{d}{=} \boldsymbol{\mu}_\star + \sqrt{\frac{\mathfrak{R}(n)}{n}}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}, \quad \mathbf{G} \sim \mathcal{N}_d\left(\mathbf{0}_d, \mathbf{I}_d\right) \perp\!\!\!\perp \mathfrak{R}(n), \tag{D.1}$$

where the random perpetuity $\mathfrak{R}(n)$ is a random variable such that

$$\mathfrak{R}(n) \overset{d}{=} 1 + \sum_{T=1}^{\infty}\left\{\prod_{t=0}^{T-1}\frac{Z_t(n)}{n}\right\}, \quad \{Z_t(n) : t \in \mathbb{Z}_+\} \overset{\text{i.i.d.}}{\sim} \chi^2(n-1). \tag{D.2}$$

In addition, we present two technical lemmas that play a key role in the proof for Theorem 4.1 below.

Lemma D.2. *Let $U \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and $V \sim \mathcal{N}\left(\mu, \tau^2\right)$ be independent Gaussian random variables. Then, their product $T := UV$ has the density function $f_T : \mathbb{R} \to \mathbb{R}_+$ such that $f_T(t) \geq f_T(-t)$ for all $t \in \mathbb{R}_+$. If we further assume that $\mu \neq 0$, then $f_T(t) > f_T(-t)$ for all $t \in (0, +\infty)$.*

Lemma D.3. *Suppose that $T_1, T_2, \cdots, T_d$ are independent random variables with density functions $f_1, f_2, \cdots, f_d$, respectively. If $f_j(t) \geq f_j(-t)$ for all $(t, j) \in \mathbb{R}_+ \times [d]$, then we have*

$$\mathbb{P}\left\{\sum_{j=1}^{d} T_j > 0\right\} \geq \frac{1}{2}. \tag{D.3}$$

Furthermore, if there exists some index $k \in [d]$ such that $f_k(t) > f_k(-t)$ for every $t \in (0, +\infty)$ and $\mathbb{P}\left\{|T_k| > \sum_{j\in[d]\setminus\{k\}} |T_j|\right\} > 0$, then

$$\mathbb{P}\left\{\sum_{j=1}^{d} T_j > 0\right\} > \frac{1}{2}. \tag{D.4}$$

See Appendices D.6 and D.7 for the proofs of Lemmas D.2 and D.3, respectively.

D.2 Proof of Theorem 4.1

Our starting point is the key characterization equation 4.3 for the limiting expected risk of the regularized plug-in LDA classifier equation 4.1:

$$\begin{aligned}
\lim_{T\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}\left\{\hat{f}_T^{\delta}(n)\right\}\right] &= \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right] \\
&= \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\Phi\left(-\frac{\boldsymbol{\mu}_\star^\top \widehat{\boldsymbol{\mu}}_\infty(n)}{\sqrt{\{\widehat{\boldsymbol{\mu}}_\infty(n)\}^\top \boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\mu}}_\infty(n)}}\right)\right] \\
&= \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathbb{P}_{\mathbf{U}\sim\mathcal{N}_d(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left\{\mathbf{U}^\top \widehat{\boldsymbol{\mu}}_\infty(n) \leq 0 \mid \widehat{\boldsymbol{\mu}}_\infty(n)\right\}\right] \\
&= \mathbb{P}_{(\mathbf{U},\mathbf{V})\sim\mathcal{N}_d(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)\otimes \mathrm{law}\{\widehat{\boldsymbol{\mu}}_\infty(n)\}}\left\{\mathbf{U}^\top \mathbf{V} \leq 0\right\},
\end{aligned} \tag{D.5}$$

where law $\{\widehat{\boldsymbol{\mu}}_\infty(n)\} \in \Delta\left(\mathbb{R}^d\right)$ refers to the (unconditional) law of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$. In view of the equation equation D.5, it is sufficient to prove that

$$\mathbb{P}_{(\mathbf{U},\mathbf{V})\sim\mathcal{N}_d(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)\otimes\text{law}\{\widehat{\boldsymbol{\mu}}_\infty(n)\}}\left\{\mathbf{U}^\top\mathbf{V}>0\right\}>\frac{1}{2} \tag{D.6}$$

in order to finish the proof for Theorem 4.1. By appealing to Lemma D.1, it follows that

$$\begin{aligned}
&\mathbb{P}_{(\mathbf{U},\mathbf{V})\sim\mathcal{N}_d(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)\otimes\text{law}\{\widehat{\boldsymbol{\mu}}_\infty(n)\}}\left\{\mathbf{U}^\top\mathbf{V}>0\right\}\\
&=\mathbb{P}_{(\mathbf{G},\mathbf{U},R)\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)\otimes\mathcal{N}_d(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)\otimes\text{law}\{\mathfrak{R}(n)\}}\left\{\mathbf{U}^\top\left(\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}\right)>0\right\}\\
&=\mathbb{P}_{(\mathbf{G},\mathbf{H},R)\sim\mathcal{N}_d^{\otimes 2}(\mathbf{0}_d,\mathbf{I}_d)\otimes\text{law}\{\mathfrak{R}(n)\}}\left\{\left(\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\boldsymbol{\mu}_\star+\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{H}\right)>0\right\},
\end{aligned} \tag{D.7}$$

where law $\{\mathfrak{R}(n)\} \in \Delta\left(\mathbb{R}_+\right)$ denotes the law of the random perpetuity equation D.2. By using the spectral decomposition $\boldsymbol{\Sigma}_\star = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$, where $\mathbf{Q} \in \mathcal{O}(d)$ and $\boldsymbol{\Lambda} = \text{diag}\left(\lambda_1, \lambda_2, \cdots, \lambda_d\right)$ for the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0$, we obtain from the equation equation D.7 that

$$\begin{aligned}
&\mathbb{P}_{(\mathbf{U},\mathbf{V})\sim\mathcal{N}_d(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)\otimes\text{law}\{\widehat{\boldsymbol{\mu}}_\infty(n)\}}\left\{\mathbf{U}^\top\mathbf{V}>0\right\}\\
&\overset{\text{(b)}}{=}\mathbb{P}_{(\mathbf{G},\mathbf{H},R)\sim\mathcal{N}_d^{\otimes 2}(\mathbf{0}_d,\mathbf{I}_d)\otimes\text{law}\{\mathfrak{R}(n)\}}\left\{\left(\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\boldsymbol{\mu}_\star+\mathbf{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)>0\right\}\\
&=\mathbb{P}_{(\mathbf{G},\mathbf{H},R)\sim\mathcal{N}_d^{\otimes 2}(\mathbf{0}_d,\mathbf{I}_d)\otimes\text{law}\{\mathfrak{R}(n)\}}\left\{\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)>0\right\},
\end{aligned} \tag{D.8}$$

where the step (b) holds due to the fact that

$$\left(\mathbf{Q}^\top\mathbf{G},\mathbf{Q}^\top\mathbf{H}\right)\overset{d}{=}(\mathbf{G},\mathbf{H})\sim\mathcal{N}_d^{\otimes 2}\left(\mathbf{0}_d,\mathbf{I}_d\right).$$

Next, we take a closer inspection on the random variable $\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)$. By simplifying as $\boldsymbol{\nu} := \mathbf{Q}^\top\boldsymbol{\mu}_\star \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$, it can be written as

$$\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)=\sum_{j=1}^d\left(\nu_j+\sqrt{\frac{R}{n}\lambda_j}G_j\right)\left(\nu_j+\sqrt{\lambda_j}H_j\right). \tag{D.9}$$

Consider $T_j(r) := \left(\nu_j+\sqrt{\frac{r}{n}\lambda_j}G_j\right)\left(\nu_j+\sqrt{\lambda_j}H_j\right)$ and its density function $f_j\left(\cdot;r\right):\mathbb{R}\to\mathbb{R}_+$ for each $(j,r)\in[d]\times(0,+\infty)$. We now fix any $r\in(0,+\infty)$. In light of Lemma D.2, one can find that $f_j\left(t;r\right)\geq f_j\left(-t;r\right)$ for every $(t,j)\in\mathbb{R}_+\times[d]$. Moreover, since $\boldsymbol{\nu}\neq\mathbf{0}_d$, there exists some index $k\in[d]$ such that $f_k\left(t;r\right)>f_k\left(-t;r\right)$ for all $t\in(0,+\infty)$. Thus, it only remains to justify the condition

$$\mathbb{P}\left\{|T_k(r)|>\sum_{j\in[d]\setminus\{k\}}|T_j(r)|\right\}>0 \tag{D.10}$$

holds for any $r\in(0,+\infty)$ in order to apply Lemma D.3. One can easily recognize that the condition equation D.10 trivially holds when $d=1$. If $d\geq 2$, then it holds that

$$\{1<T_k(r)<2\}\cap\bigcap_{j\in[d]\setminus\{k\}}\left\{0<T_j(r)<\frac{1}{2d}\right\}\subseteq\left\{|T_k(r)|>\sum_{j\in[d]\setminus\{k\}}|T_j(r)|\right\}. \tag{D.11}$$

Since $T_j(r)$ assigns strictly positive probability to every non-empty interval on the real line $\mathbb{R}$ for any $j\in[d]$, the desired condition equation D.10 immediately follows from the relation equation D.11.

Finally, we are ready to finish the proof of Theorem 4.1 by applying Lemma D.3 to the i.i.d. random variables $\{T_j(r) : j \in [d]\}$ to obtain

$$\begin{aligned}&\mathbb{P}_{(\mathbf{G},\mathbf{H})\sim\mathcal{N}_d^{\otimes 2}(\mathbf{0}_d,\mathbf{I}_d)}\left\{\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\sqrt{\frac{r}{n}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)>0\right\}\\ &=\mathbb{P}\left\{\sum_{j=1}^d T_j(r)>0\right\}>\frac{1}{2}\end{aligned} \tag{D.12}$$

for every fixed $r \in (0, +\infty)$. Hence, it follows from the equation equation D.8 and the inequality equation D.12 that

$$\begin{aligned}&\mathbb{P}_{(\mathbf{U},\mathbf{V})\sim\mathcal{N}_d(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)\otimes\text{law}\{\widehat{\boldsymbol{\mu}}_\infty(n)\}}\left\{\mathbf{U}^\top\mathbf{V}>0\right\}\\ &=\mathbb{P}_{(\mathbf{G},\mathbf{H},R)\sim\mathcal{N}_d^{\otimes 2}(\mathbf{0}_d,\mathbf{I}_d)\otimes\text{law}\{\mathfrak{R}(n)\}}\left\{\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)>0\right\}\\ &=\mathbb{E}_{R\sim\text{law}\{\mathfrak{R}(n)\}}\left[\mathbb{P}_{(\mathbf{G},\mathbf{H})\sim\mathcal{N}_d^{\otimes 2}(\mathbf{0}_d,\mathbf{I}_d)}\left\{\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\sqrt{\frac{R}{n}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{G}\right)^\top\left(\mathbf{Q}^\top\boldsymbol{\mu}_\star+\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{H}\right)>0\right\}\right]\\ &>\frac{1}{2},\end{aligned}$$

concluding the proof for the ultimate claim equation D.6 as well as Theorem 4.1.

D.3 Proof of Theorem 4.2

The proof of the Theorem 4.2 is a straightforward consequence of the following key lemma together with the classical Mills ratio upper bound:

$$\Phi(-x)\leq\frac{1}{x\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),\quad x\in(0,+\infty).$$

Lemma D.4. *Given any ground-truth parameter $(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \in \Theta$ and any fixed sample size $n \geq d+1$, it holds that*

$$0\leq\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right]-\mathcal{R}_\star\leq\Phi\left(-\frac{\Delta_\star}{3\kappa(\boldsymbol{\Sigma}_\star)}\right)+\frac{4d\kappa(\boldsymbol{\Sigma}_\star)^2}{\Delta_\star^2}. \tag{D.13}$$

Proof of Lemma D.4. In order to prove Lemma D.4, we first need the following technical result.

Lemma D.5. *If $n \geq d + 1$, then the expectation of the random perpetuity equation D.1 is given by*

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathfrak{R}(n)\right]=n. \tag{D.14}$$

Proof of Lemma D.5. Since all the summands in the definition equation D.1 of the random perpetuity $\mathfrak{R}(n)$ are non-negative, Tonelli's theorem yields

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathfrak{R}(n)\right]=1+\sum_{T=1}^\infty\mathbb{E}\left[\prod_{t=0}^{T-1}\frac{Z_t(n)}{n}\right] \tag{D.15}$$

Since $\{Z_t(n) : t \in \mathbb{Z}_+\} \overset{\text{i.i.d.}}{\sim} \chi^2(n-1)$, one can see that

$$\mathbb{E}\left[\prod_{t=0}^{T-1}\frac{Z_t(n)}{n}\right]=\left(1-\frac{1}{n}\right)^T,\quad T\in\mathbb{N}. \tag{D.16}$$

Taking the equation equation D.16 collectively into the identity equation D.15 then deduces

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathfrak{R}(n)\right]=1+\sum_{T=1}^\infty\left(1-\frac{1}{n}\right)^T=\sum_{T=0}^\infty\left(1-\frac{1}{n}\right)^T=n,$$

as desired.

□

To begin with, we introduce

$$\boldsymbol{\beta}_\star := \boldsymbol{\Sigma}_\star^{-\frac{1}{2}} \boldsymbol{\mu}_\star \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_\infty(n) := \boldsymbol{\Sigma}_\star^{-\frac{1}{2}} \widehat{\boldsymbol{\mu}}_\infty(n).$$

Then, it is clear that $\|\boldsymbol{\beta}_\star\|_2 = \Delta_\star$, and we obtain by Lemma D.1 that

$$\widehat{\boldsymbol{\beta}}_\infty(n) = \boldsymbol{\beta}_\star + \sqrt{\frac{\mathfrak{R}(n)}{n}}\mathbf{G}, \quad \mathbf{G} \sim \mathcal{N}_d\left(\mathbf{0}_d, \mathbf{I}_d\right) \perp\!\!\!\perp \mathfrak{R}(n), \tag{D.17}$$

where $\mathfrak{R}(n)$ is defined as equation D.2. Here, let us consider

$$M_\infty(n) := \frac{\boldsymbol{\mu}_\star^\top \widehat{\boldsymbol{\mu}}_\infty(n)}{\sqrt{\{\widehat{\boldsymbol{\mu}}_\infty(n)\}^\top \boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\mu}}_\infty(n)}} = \frac{\boldsymbol{\beta}_\star^\top \boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\beta}}_\infty(n)}{\left\|\boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\beta}}_\infty(n)\right\|_2}, \tag{D.18}$$

so that $\mathcal{R}_\infty(n) = \Phi\{-M_\infty(n)\}$. To further control the limiting expected risk $\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right]$ of the regularized plug-in LDA classifier equation 4.1, we define the event

$$\mathcal{E}_\eta := \left\{\left\|\widehat{\boldsymbol{\beta}}_\infty(n) - \boldsymbol{\beta}_\star\right\|_2 \leq \eta\Delta_\star\right\}, \quad \eta \in (0,1).$$

On the event $\mathcal{E}_\eta$, it holds that

$$\begin{aligned}
\boldsymbol{\beta}_\star^\top \boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\beta}}_\infty(n) &= \boldsymbol{\beta}_\star^\top \boldsymbol{\Sigma}_\star \boldsymbol{\beta}_\star + \boldsymbol{\beta}_\star^\top \boldsymbol{\Sigma}_\star (\widehat{\boldsymbol{\beta}}_\infty(n) - \boldsymbol{\beta}_\star) \\
&\overset{(a)}{\geq} \lambda_{\min}(\boldsymbol{\Sigma}_\star)\Delta_\star^2 + \boldsymbol{\beta}_\star^\top \boldsymbol{\Sigma}_\star (\widehat{\boldsymbol{\beta}}_\infty(n) - \boldsymbol{\beta}_\star) \\
&\overset{(b)}{\geq} \lambda_{\min}(\boldsymbol{\Sigma}_\star)\Delta_\star^2 - \|\boldsymbol{\Sigma}_\star\|_{\mathrm{op}} \|\boldsymbol{\beta}_\star\|_2 \|\widehat{\boldsymbol{\beta}}_\infty(n) - \boldsymbol{\beta}_\star\|_2 \\
&\overset{(c)}{\geq} \lambda_{\min}(\boldsymbol{\Sigma}_\star)\Delta_\star^2 - (\lambda_{\max}(\boldsymbol{\Sigma}_\star)\Delta_\star) \cdot (\eta\Delta_\star) \\
&= (\lambda_{\min}(\boldsymbol{\Sigma}_\star) - \lambda_{\max}(\boldsymbol{\Sigma}_\star)\eta)\Delta_\star^2,
\end{aligned} \tag{D.19}$$

where the steps (a) and (c) is due to the spectral property of the $\boldsymbol{\Sigma}_\star$ and (b) is due to the Cauchy-Schwarz inequality. Let

$$\eta = \frac{\lambda_{\min}(\boldsymbol{\Sigma}_\star)}{2\lambda_{\max}(\boldsymbol{\Sigma}_\star)} = \frac{1}{2\kappa(\boldsymbol{\Sigma}_\star)}. \tag{D.20}$$

Thereby, combining equation D.19 and equation D.20 yield

$$\boldsymbol{\beta}_\star^\top \boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\beta}}_\infty(n) \geq \frac{\lambda_{\min}(\boldsymbol{\Sigma}_\star)}{2}\Delta_\star^2. \tag{D.21}$$

Using the equation D.20 on the event $\mathcal{E}_\eta$, the denominator can be controlled by

$$\begin{aligned}
\|\boldsymbol{\Sigma}_\star \widehat{\boldsymbol{\beta}}_\infty(n)\|_2 &= \sqrt{\widehat{\boldsymbol{\beta}}_\infty(n)^\top \boldsymbol{\Sigma}_\star^2 \widehat{\boldsymbol{\beta}}_\infty(n)} \\
&\overset{(a)}{\leq} \lambda_{\max}(\boldsymbol{\Sigma}_\star)\|\widehat{\boldsymbol{\beta}}_\infty(n)\|_2 \\
&\leq \lambda_{\max}(\boldsymbol{\Sigma}_\star)(\|\boldsymbol{\beta}_\star\|_2 + \eta\Delta_\star) \\
&= \lambda_{\max}(\boldsymbol{\Sigma}_\star)(1+\eta)\Delta_\star \\
&\leq \frac{3}{2}\lambda_{\max}(\boldsymbol{\Sigma}_\star)\Delta_\star,
\end{aligned} \tag{D.22}$$

where the step (a) is due to the spectral property of $\boldsymbol{\Sigma}_\star$. Combining two bounds equation D.21 and equation D.22 together, we obtain

$$\begin{aligned}
\mathcal{R}_\infty(n) \cdot \mathbb{1}_{\mathcal{E}_\eta} &= \Phi\{-M_\infty(n)\} \cdot \mathbb{1}_{\mathcal{E}_\eta} \\
&\leq \Phi\left(-\frac{(\lambda_{\min}(\Delta_\star)/2)}{(3\lambda_{\max}(\Delta_\star)/2)}\Delta_\star\right)\mathbb{1}_{\mathcal{E}_\eta} \\
&\leq \Phi\left(-\frac{\Delta_\star}{3\kappa(\boldsymbol{\Sigma}_\star)}\right).
\end{aligned} \tag{D.23}$$

Taking expectation to both sides of the inequality equation D.23, one can reveal that

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right] &= \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\cdot\mathbb{1}_{\mathcal{E}_\eta}\right] + \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\cdot\mathbb{1}_{\mathcal{E}_\eta^c}\right] \\ &\overset{\text{(a)}}{\leq} \Phi\left(-\frac{\Delta_\star}{3\kappa(\boldsymbol{\Sigma}_\star)}\right) + \mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left\{\mathcal{E}_\eta^c\right\},\end{aligned} \tag{D.24}$$

where the step (a) makes use of the trivial bound $\Phi(\cdot)\leq 1$. Leveraging Markov's inequality together with Lemma D.5 then yields

$$\mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left\{\mathcal{E}_\eta^c\right\} = \mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left\{\left\|\widehat{\boldsymbol{\beta}}_\infty(n)-\boldsymbol{\beta}_\star\right\|_2 > \eta\Delta_\star\right\} \leq \frac{\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\left\|\widehat{\boldsymbol{\beta}}_\infty(n)-\boldsymbol{\beta}_\star\right\|_2^2\right]}{\eta^2\Delta_\star^2}. \tag{D.25}$$

Now, it only remains to compute the expectation $\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\left\|\widehat{\boldsymbol{\beta}}_\infty(n)-\boldsymbol{\beta}_\star\right\|_2^2\right]$. Appealing to Lemma D.1, we obtain the following distributional representation of $\widehat{\boldsymbol{\beta}}_\infty(n)$:

$$\widehat{\boldsymbol{\beta}}_\infty(n) \overset{d}{=} \boldsymbol{\beta}_\star + \sqrt{\frac{\mathfrak{R}(n)}{n}}\mathbf{G}, \quad \mathbf{G}\sim\mathcal{N}_d\left(\mathbf{0}_d,\mathbf{I}_d\right) \perp\!\!\!\perp \mathfrak{R}(n).$$

Thus, the expectation $\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\left\|\widehat{\boldsymbol{\beta}}_\infty(n)-\boldsymbol{\beta}_\star\right\|_2^2\right]$ can be computed as

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\left\|\widehat{\boldsymbol{\beta}}_\infty(n)-\boldsymbol{\beta}_\star\right\|_2^2\right] = \frac{1}{n}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\|\mathbf{G}\|_2^2\cdot\mathfrak{R}(n)\right] \overset{\text{(d)}}{=} d, \tag{D.26}$$

where the step (d) is a result of the independence between $\mathbf{G}\sim\mathcal{N}_d\left(\mathbf{0}_d,\mathbf{I}_d\right)$ and $\mathfrak{R}(n)$ together with Lemma D.5. Taking the equation equation D.26 into the bound equation D.25 and then substituting it into the main inequality equation D.24, we eventually reach

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}_\infty(n)\right] \leq \Phi\left(-\frac{\Delta_\star}{3\kappa(\boldsymbol{\Sigma}_\star)}\right) + \frac{d}{\eta^2\Delta_\star^2}. \tag{D.27}$$

Hence, taking $\eta = \frac{1}{2\kappa(\boldsymbol{\Sigma}_\star)}$ in the final bound equation D.27 concludes the proof of Lemma D.4.

□

D.4 Proof of Theorem 4.3

Similar to the proof of Proposition 3.1, we observe that if $\mathbf{O}_{1:n}^{(t)} = \left(O_i^{(t)} = \left(\mathbf{X}_i^{(t)}, Y_i^{(t)}\right) : i\in[n]\right) \sim \mathbb{P}_{(\widehat{\boldsymbol{\mu}}_t^\star(n),\boldsymbol{\Sigma}_\star)}^{\otimes n}$, then we have

$$\begin{aligned}\mathbf{Y}_{1:n}^{(t)} &= \left(Y_i^{(t)} : i\in[n]\right) \sim \text{Unif}\left(\{\pm 1\}^n\right), \\ \mathbf{X}_i^{(t)} &= Y_i^{(t)}\hat{\boldsymbol{\mu}}_t^\star(n) + \boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{Z}_i^{(t)}, \quad \forall i\in[n],\end{aligned} \tag{D.28}$$

where $\left(\mathbf{Z}_i^{(t)} : i\in[n]\right) \sim \left\{\mathcal{N}_d\left(\mathbf{0}_d,\mathbf{I}_d\right)\right\}^{\otimes n}$. The observation equation D.28 then yields

$$\widehat{\boldsymbol{\mu}}_{t+1}^\star(n) = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^{(t)}Y_i^{(t)} = \widehat{\boldsymbol{\mu}}_t^\star(n) + \frac{1}{n}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\sum_{i=1}^n Y_i^{(t)}\mathbf{Z}_i^{(t)} \tag{D.29}$$

for every $t\in\mathbb{Z}_+$. Taking the summation of the equation equation D.29 over $t=0,1,\cdots,T-1$, we reach

$$\widehat{\boldsymbol{\mu}}_T^\star(n) = \boldsymbol{\mu}_\star + \boldsymbol{\Sigma}_\star^{\frac{1}{2}}\left[\sum_{t=0}^{T-1}\left\{\frac{1}{n}\sum_{i=1}^n Y_i^{(t)}\mathbf{Z}_i^{(t)}\right\}\right] \overset{\text{(a)}}{\sim} \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{T}{n}\boldsymbol{\Sigma}_\star\right), \tag{D.30}$$

where the step (a) holds since $\left\{\frac{1}{n}\sum_{i=1}^n Y_i^{(t)}\mathbf{Z}_i^{(t)} : t\in\mathbb{Z}_+\right\} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d\left(\mathbf{0}_d, \frac{1}{n}\mathbf{I}_d\right)$. Thus, we may write

$$\widehat{\boldsymbol{\mu}}_T^\star(n) \overset{d}{=} \boldsymbol{\mu}_\star + \sqrt{\frac{T}{n}}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}, \quad \mathbf{G}\sim\mathcal{N}_d\left(\mathbf{0}_d,\mathbf{I}_d\right). \tag{D.31}$$

By appealing to Proposition 2.1, the expected classification risk of the induced plug-in LDA classifier can be simplified as

$$\begin{aligned}
&\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}\left\{\hat{f}_T^\star(n)\right\}\right] \\
&= \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\mathcal{R}\left\{f_{\widehat{\mathbf{w}}_T^\star(n)}\right\}\right] \\
&= \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\Phi\left\{-\frac{\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-1}\hat{\boldsymbol{\mu}}_T^\star(n)}{\sqrt{\{\hat{\boldsymbol{\mu}}_T^\star(n)\}^\top\boldsymbol{\Sigma}_\star^{-1}\hat{\boldsymbol{\mu}}_T^\star(n)}}\right\}\right] \\
&\overset{\text{(b)}}{=} \mathbb{E}_{\mathbf{G}\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)}\left[\Phi\left\{-\frac{\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-1}\boldsymbol{\mu}_\star+\sqrt{\frac{T}{n}}\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\mathbf{G}}{\sqrt{\left\{\boldsymbol{\mu}_\star+\sqrt{\frac{T}{n}}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}\right\}^\top\boldsymbol{\Sigma}_\star^{-1}\left\{\boldsymbol{\mu}_\star+\sqrt{\frac{T}{n}}\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}\right\}}}\right\}\right] \\
&= \mathbb{E}_{\mathbf{G}\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)}\left[\Phi\left\{-\frac{\sqrt{\frac{n}{T}}\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-1}\boldsymbol{\mu}_\star+\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\mathbf{G}}{\sqrt{\left\{\sqrt{\frac{n}{T}}\boldsymbol{\mu}_\star+\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}\right\}^\top\boldsymbol{\Sigma}_\star^{-1}\left\{\sqrt{\frac{n}{T}}\boldsymbol{\mu}_\star+\boldsymbol{\Sigma}_\star^{\frac{1}{2}}\mathbf{G}\right\}}}\right\}\right] \\
&\overset{\text{(c)}}{\to} \mathbb{E}_{\mathbf{G}\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)}\left[\Phi\left\{-\frac{\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\mathbf{G}}{\|\mathbf{G}\|_2}\right\}\right] \\
&\overset{\text{(d)}}{=} \frac{1}{2}
\end{aligned} \tag{D.32}$$

as $T\to\infty$, where the step (b) invokes the fact equation D.31, the step (c) is due to the bounded convergence theorem, and the step (d) holds since

$$\begin{aligned}
\mathbb{E}_{\mathbf{G}\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)}\left[\Phi\left\{-\frac{\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\mathbf{G}}{\|\mathbf{G}\|_2}\right\}\right] &\overset{\text{(e)}}{=} \mathbb{E}_{\mathbf{G}\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)}\left[\Phi\left\{\frac{\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\mathbf{G}}{\|\mathbf{G}\|_2}\right\}\right] \\
&= 1-\mathbb{E}_{\mathbf{G}\sim\mathcal{N}_d(\mathbf{0}_d,\mathbf{I}_d)}\left[\Phi\left\{-\frac{\boldsymbol{\mu}_\star^\top\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\mathbf{G}}{\|\mathbf{G}\|_2}\right\}\right].
\end{aligned}$$

Here, the step (e) results from the fact that $\mathbf{G}\overset{d}{=}-\mathbf{G}\sim\mathcal{N}_d\left(\mathbf{0}_d,\mathbf{I}_d\right)$, and this concludes the proof of Theorem 4.3.

D.5 Proof of Lemma D.1

To begin with, we define the *covariance-path σ-field* defined on $\mathbb{R}^d\times\{\pm1\}$:

$$\mathcal{V}(n):=\sigma\left(\left\{\widehat{\boldsymbol{\Sigma}}_t(n):t\in\mathbb{Z}_+\right\}\right).$$

The first step is concerned with the conditional distribution of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ given the covariance-path σ-field $\mathcal{V}(n)$, which is accomplished through the following lemma.

Lemma D.6. *If $n\geq 2$, then the following properties hold:*

(i) The matrix infinite sum $\sum_{T=0}^\infty\widehat{\boldsymbol{\Sigma}}_T(n)$ converges $\mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}$-almost surely to a $d\times d$ symmetric positive definite matrix.

(ii) The conditional distribution of the limiting point $\widehat{\boldsymbol{\mu}}_\infty(n)$ given the covariance-path σ-field $\mathcal{V}(n)$ is given by

$$\widehat{\boldsymbol{\mu}}_\infty(n)|\,\mathcal{V}(n)\sim\mathcal{N}_d\left(\boldsymbol{\mu}_\star,\frac{1}{n}\sum_{T=0}^\infty\widehat{\boldsymbol{\Sigma}}_T(n)\right). \tag{D.33}$$

Proof of Lemma D.6.

(i) It can be easily observed that

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\text{Trace}\left\{\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right\}\right] &= \sum_{T=0}^{\infty}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\text{Trace}\left\{\widehat{\boldsymbol{\Sigma}}_T(n)\right\}\right] \\ &\overset{\text{(a)}}{=} \sum_{T=0}^{\infty}\left(1-\frac{1}{n}\right)^T \text{Trace}\left(\boldsymbol{\Sigma}_\star\right) \\ &= n\cdot\text{Trace}\left(\boldsymbol{\Sigma}_\star\right) \\ &< +\infty,\end{aligned}$$

where the step (a) leverages the conditional contraction identity equation 3.3. Thus, it follows that

$$\text{Trace}\left\{\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right\} < +\infty \quad \mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\text{-almost surely,}$$

which in turn implies the $\mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}$-almost sure convergence of the matrix series $\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)$.

(ii) The consequence equation 3.2 of Proposition 3.1 tells us that

$$\widehat{\boldsymbol{\mu}}_T(n)|\,\mathcal{V}(n) \sim \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{1}{n}\sum_{t=0}^{T-1}\widehat{\boldsymbol{\Sigma}}_t(n)\right). \tag{D.34}$$

Thus, the conditional characteristic function of $\widehat{\boldsymbol{\mu}}_T(n)$ given the covariance-path σ-field $\mathcal{V}(n)$ equals to

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_T(n)\right\}\middle|\,\mathcal{V}(n)\right] = \exp\left[i\mathbf{v}^\top\boldsymbol{\mu}_\star - \frac{1}{2n}\mathbf{v}^\top\left\{\sum_{t=0}^{T-1}\widehat{\boldsymbol{\Sigma}}_t(n)\right\}\mathbf{v}\right]. \tag{D.35}$$

By virtue of Theorem 3.1, letting $T\to\infty$ to both sides of equation D.35 yields

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_\infty(n)\right\}\middle|\,\mathcal{V}(n)\right] &= \lim_{T\to\infty}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_T(n)\right\}\middle|\,\mathcal{V}(n)\right] \\ &= \lim_{T\to\infty}\exp\left[i\mathbf{v}^\top\boldsymbol{\mu}_\star - \frac{1}{2n}\mathbf{v}^\top\left\{\sum_{t=0}^{T-1}\widehat{\boldsymbol{\Sigma}}_t(n)\right\}\mathbf{v}\right] \\ &= \exp\left[i\mathbf{v}^\top\boldsymbol{\mu}_\star - \frac{1}{2n}\mathbf{v}^\top\left\{\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right\}\mathbf{v}\right],\end{aligned} \tag{D.36}$$

which concludes that

$$\widehat{\boldsymbol{\mu}}_\infty(n)|\,\mathcal{V}(n) \sim \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{1}{n}\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right)$$

as desired.

□

For notational simplicity, let $\mathbb{V}_\infty(n) := \frac{1}{n}\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)$. Then, the (unconditional) characteristic function of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ can be written as

$$\begin{aligned}\text{ch}_{\widehat{\boldsymbol{\mu}}_\infty(n)}(\mathbf{v}) &:= \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_\infty(n)\right\}\right] \\ &= \exp\left(i\mathbf{v}^\top\boldsymbol{\mu}_\star\right)\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{-\frac{1}{2}\mathbf{v}^\top\mathbb{V}_\infty(n)\mathbf{v}\right\}\right].\end{aligned} \tag{D.37}$$

In view of the equation equation D.37, we next take a closer look at the marginal law of $\mathbf{v}^\top\mathbb{V}_\infty(n)\mathbf{v}$. In order to do so, let

$$\overline{\boldsymbol{\Sigma}}_t(n) := \boldsymbol{\Sigma}_\star^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_t(n)\boldsymbol{\Sigma}_\star^{-\frac{1}{2}}, \quad t\in\mathbb{Z}_+,$$

and $\overline{\mathbb{V}}_\infty(n) := \boldsymbol{\Sigma}_\star^{-\frac{1}{2}} \mathbb{V}_\infty(n) \boldsymbol{\Sigma}_\star^{-\frac{1}{2}} = \frac{1}{n} \sum_{T=0}^{\infty} \overline{\boldsymbol{\Sigma}}_T(n)$. Let us claim that $\mathbf{R}\overline{\mathbb{V}}_\infty(n)\mathbf{R}^\top \overset{d}{=} \overline{\mathbb{V}}_\infty(n)$ for every $\mathbf{R} \in \mathcal{O}(d)$, where

$$\mathcal{O}(d) := \left\{ \mathbf{R} \in \mathbb{R}^{d \times d} : \mathbf{R}^\top \mathbf{R} = \mathbf{R}\mathbf{R}^\top = \mathbf{I}_d \right\}$$

denotes the orthogonal group in dimension d. By the part (ii) of Proposition 3.1, we have

$$n \cdot \overline{\boldsymbol{\Sigma}}_{T+1}(n) \big| \, \mathcal{F}_T(n) \sim \mathbb{W}_d \left(n - 1, \overline{\boldsymbol{\Sigma}}_T(n) \right), \quad T \in \mathbb{Z}_+. \tag{D.38}$$

Thus, we may write $\overline{\boldsymbol{\Sigma}}_0(n) = \mathbf{I}_d$ and

$$\overline{\boldsymbol{\Sigma}}_{T+1}(n) = \frac{1}{n} \overline{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}(n) \mathbf{W}_T(n) \overline{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}(n), \quad T \in \mathbb{Z}_+, \tag{D.39}$$

where $\{\mathbf{W}_t(n) : t \in \mathbb{Z}_+\} \overset{\text{i.i.d.}}{\sim} \mathbb{W}_d(n-1, \mathbf{I}_d)$. Due to the expression equation D.39, one can readily observe that $\overline{\boldsymbol{\Sigma}}_{T+1}(n)$ is a function of $(\mathbf{W}_0(n), \mathbf{W}_1(n), \cdots, \mathbf{W}_T(n))$ for each $T \in \mathbb{Z}_+$. Therefore, one can write

$$\overline{\mathbb{V}}_\infty(n) = \Psi \left(\mathbf{W}_t(n) : t \in \mathbb{Z}_+ \right) \tag{D.40}$$

for some function $\Psi : \left(\mathcal{S}_+^d \right)^\infty \to \mathcal{S}_+^d$. Also, one can reveal that

$$\mathbf{R}\overline{\boldsymbol{\Sigma}}_{T+1}(n)\mathbf{R}^\top = \frac{1}{n} \left\{ \mathbf{R}\overline{\boldsymbol{\Sigma}}_T(n)\mathbf{R}^\top \right\}^{\frac{1}{2}} \mathbf{R}\mathbf{W}_T(n)\mathbf{R}^\top \left\{ \mathbf{R}\overline{\boldsymbol{\Sigma}}_T(n)\mathbf{R}^\top \right\}^{\frac{1}{2}}, \quad T \in \mathbb{Z}_+$$

for any $\mathbf{R} \in \mathcal{O}(d)$, thereby

$$\mathbf{R}\overline{\mathbb{V}}_\infty(n)\mathbf{R}^\top = \Psi \left(\mathbf{R}\mathbf{W}_t(n)\mathbf{R}^\top : t \in \mathbb{Z}_+ \right). \tag{D.41}$$

Taking two observations equation D.40 and equation D.41 collectively establishes the desired claim $\mathbf{R}\overline{\mathbb{V}}_\infty(n)\mathbf{R}^\top \overset{d}{=} \overline{\mathbb{V}}_\infty(n)$ for every $\mathbf{R} \in \mathcal{O}(d)$, since

$$\left(\mathbf{W}_t(n) : t \in \mathbb{Z}_+ \right) \overset{d}{=} \left(\mathbf{R}\mathbf{W}_t(n)\mathbf{R}^\top : t \in \mathbb{Z}_+ \right), \quad \mathbf{R} \in \mathcal{O}(d).$$

Thus, one can reach

$$\mathbf{u}^\top \overline{\mathbb{V}}_\infty(n) \mathbf{u} \overset{d}{=} (\mathbf{R}\mathbf{u})^\top \overline{\mathbb{V}}_\infty(n) (\mathbf{R}\mathbf{u}), \quad (\mathbf{u}, \mathbf{R}) \in \left(\mathbb{R}^d \setminus \{\mathbf{0}_d\} \right) \times \mathcal{O}(d). \tag{D.42}$$

In particular, by taking a proper $d \times d$ orthogonal matrix $\mathbf{R} \in \mathcal{O}(d)$ so that $\mathbf{R} \left(\frac{\mathbf{u}}{\|\mathbf{u}\|_2} \right) = \mathbf{e}_1$ into the equation equation D.42, we deduce

$$\mathbf{u}^\top \overline{\mathbb{V}}_\infty(n) \mathbf{u} \overset{d}{=} \|\mathbf{u}\|_2^2 \cdot \mathbf{e}_1^\top \overline{\mathbb{V}}_\infty(n) \mathbf{e}_1, \quad \mathbf{u} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}. \tag{D.43}$$

Letting $\mathbf{u} = \boldsymbol{\Sigma}_\star^{\frac{1}{2}} \mathbf{v}$ in the conclusion equation D.43 then yields

$$\mathbf{v}^\top \mathbb{V}_\infty(n) \mathbf{v} \overset{d}{=} \left(\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v} \right) \cdot \mathbf{e}_1^\top \overline{\mathbb{V}}_\infty(n) \mathbf{e}_1. \tag{D.44}$$

Now, it only remains to study the random variable $\mathbf{e}_1^\top \overline{\mathbb{V}}_\infty(n) \mathbf{e}_1$. Define $Z_t(n) := n \cdot \frac{\mathbf{e}_1^\top \overline{\boldsymbol{\Sigma}}_{t+1}(n) \mathbf{e}_1}{\mathbf{e}_1^\top \overline{\boldsymbol{\Sigma}}_t(n) \mathbf{e}_1}$ for $t \in \mathbb{Z}_+$. The finding equation D.38 tells us that

$$\{Z_t(n) : t \in \mathbb{Z}_+\} \overset{\text{i.i.d.}}{\sim} \chi^2(n-1).$$

Therefore, one may conclude that

$$\mathbf{e}_1^\top \overline{\mathbb{V}}_\infty(n) \mathbf{e}_1 = \frac{1}{n} \sum_{T=0}^{\infty} \mathbf{e}_1^\top \overline{\boldsymbol{\Sigma}}_T(n) \mathbf{e}_1 = \frac{1}{n} \left[1 + \sum_{T=1}^{\infty} \left\{ \prod_{t=0}^{T-1} \frac{Z_t(n)}{n} \right\} \right] \overset{d}{=} \frac{\mathfrak{R}(n)}{n}. \tag{D.45}$$

Taking two facts equation D.44 and equation D.45 collectively into equation D.37, the (unconditional) characteristic function for the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ becomes

$$\mathrm{ch}_{\widehat{\boldsymbol{\mu}}_\infty(n)}(\mathbf{v}) = \exp\left(i\mathbf{v}^\top \boldsymbol{\mu}_\star \right) \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\exp\left\{ -\frac{1}{2} \mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v} \cdot \frac{\mathfrak{R}(n)}{n} \right\} \right], \tag{D.46}$$

thereby concluding the proof of Lemma D.1, since the right-hand side of the equation equation D.46 exactly coincides with the characteristic function of $\boldsymbol{\mu}_\star + \sqrt{\frac{\mathfrak{R}(n)}{n}} \boldsymbol{\Sigma}_\star^{\frac{1}{2}} \mathbf{G}$, where $\mathbf{G} \sim \mathcal{N}_d(\mathbf{0}_d, \mathbf{I}_d) \perp\!\!\!\perp \mathfrak{R}(n)$.

D.6 Proof of Lemma D.2

The cumulative distribution function of $T = UV$ can be expressed as

$$\begin{aligned}\mathbb{P}\{T \le t\} &= \mathbb{P}\{UV \le t\} \\ &= \int_{\mathbb{R}} \mathbb{P}\{UV \le t \mid U = u\} f_U(u)\mathrm{d}u \\ &= \int_0^{\infty} \mathbb{P}\left\{V \le \frac{t}{u}\right\} f_U(u)\mathrm{d}u + \int_{-\infty}^{0} \mathbb{P}\left\{V \ge \frac{t}{u}\right\} f_U(u)\mathrm{d}u.\end{aligned} \tag{D.47}$$

By differentiating both sides of the equation equation D.47 with respect to t, the density function of $T = UV$ can be simplified as

$$f_T(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\{T \le t\} = \int_{\mathbb{R}} f_U(u) f_V\left(\frac{t}{u}\right) \cdot \frac{1}{|u|}\mathrm{d}u, \quad t \in \mathbb{R}. \tag{D.48}$$

Thus, one can reveal for any $t \in \mathbb{R}_+$ that

$$\begin{aligned}f_T(t) - f_T(-t) &= \int_{\mathbb{R}} \frac{f_U(u)}{|u|}\left\{f_V\left(\frac{t}{u}\right) - f_V\left(-\frac{t}{u}\right)\right\}\mathrm{d}u \\ &= \int_0^{\infty} \frac{1}{u}\left\{f_U(u) - f_U(-u)\right\}\left\{f_V\left(\frac{t}{u}\right) - f_V\left(-\frac{t}{u}\right)\right\}\mathrm{d}u \\ &= \int_0^{\infty} \frac{f_U(-u) f_V\left(-\frac{t}{u}\right)}{u}\left\{\exp\left(\frac{2\mu u}{\sigma^2}\right) - 1\right\}\left\{\exp\left(\frac{2\mu t}{\tau^2 u}\right) - 1\right\}\mathrm{d}u \\ &\overset{\text{(a)}}{\ge} 0,\end{aligned} \tag{D.49}$$

where the step (a) follows since $\left\{\exp\left(\frac{2\mu u}{\sigma^2}\right) - 1\right\}\left\{\exp\left(\frac{2\mu t}{\tau^2 u}\right) - 1\right\} \ge 0$ for all $u \in \mathbb{R}_+$. Moreover, the remaining conclusion follows when $\mu \ne 0$ since the integrand is strictly positive for $u \in (0, +\infty)$, concluding the proof for Lemma D.2.

D.7 Proof of Lemma D.3

Firstly, let us define

$$R_j := |T_j| \quad \text{and} \quad \varepsilon_j := \operatorname{sign}(T_j)$$

for each $j \in [d]$. Then, one can easily find that

$$p_j(r) := \mathbb{P}\{\varepsilon_j = +1 \mid R_j = r\} = \frac{f_j(r)}{f_j(r) + f_j(-r)} \overset{\text{(a)}}{\ge} \frac{1}{2}, \quad (r, j) \in (0, +\infty) \times [d], \tag{D.50}$$

where the step (a) follows by the condition for which $f_j(t) \ge f_j(-t)$ for all $(t, j) \in \mathbb{R}_+ \times [d]$. Here, we introduce

$$\mathcal{A}(\mathbf{r}_{1:d}) := \left\{\boldsymbol{\sigma}_{1:d} \in \{\pm 1\}^d : \sum_{j=1}^{d} \sigma_j r_j > 0\right\}, \quad \mathbf{r}_{1:d} \in (0, +\infty)^d,$$

which turns out to be a *non-decreasing subset* of $\{\pm 1\}^d$, i.e., if $\boldsymbol{\sigma}_{1:d} \in \mathcal{A}(\mathbf{r}_{1:d})$ and $\boldsymbol{\tau}_{1:d} \in \{\pm 1\}^d$ satisfies $\tau_j \ge \sigma_j$ for every $j \in [d]$, then $\boldsymbol{\tau}_{1:d} \in \mathcal{A}(\mathbf{r}_{1:d})$.

Claim D.1. *If $\boldsymbol{\eta}_{1:d} \sim \mathrm{Unif}\left(\{\pm 1\}^d\right) \perp\!\!\!\perp \mathbf{T}_{1:d}$, then $\boldsymbol{\varepsilon}_{1:d}$ stochastically dominates $\boldsymbol{\eta}_{1:d}$ on $\{\pm 1\}^d$. That is, for any non-decreasing subset $A \subseteq \{\pm 1\}^d$, we have*

$$\mathbb{P}\{\boldsymbol{\varepsilon}_{1:d} \in A \mid \mathbf{R}_{1:d} = \mathbf{r}_{1:d}\} \ge \mathbb{P}\{\boldsymbol{\eta}_{1:d} \in A\}, \quad \mathbf{r}_{1:d} \in (0, +\infty)^d. \tag{D.51}$$

Proof of Claim D.1. Before proceeding, let $U_1, U_2, \cdots, U_d \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0, 1)$, and $\boldsymbol{\alpha}(\mathbf{r}_{1:d}) : (0, 1)^d \to \{\pm 1\}^d$ and $\boldsymbol{\beta} : (0, 1)^d \to \{\pm 1\}^d$ be functions defined by

$$\alpha_j(\mathbf{r}_{1:d})(\mathbf{u}_{1:d}) := \begin{cases} +1 & \text{if } u_j \le p_j(r_j); \\ -1 & \text{otherwise,} \end{cases} \qquad \beta_j(\mathbf{u}_{1:d}) := \begin{cases} +1 & \text{if } u_j \le \frac{1}{2}; \\ -1 & \text{otherwise,} \end{cases}$$

for each $j \in [d]$. Then, it is straightforward to find that

$$\begin{aligned} \mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in A \mid \mathbf{R}_{1:d} = \mathbf{r}_{1:d}\right\} &= \mathbb{P}\left\{\boldsymbol{\alpha}\left(\mathbf{r}_{1:d}\right)\left(\mathbf{U}_{1:d}\right) \in A\right\}, \\ \mathbb{P}\left\{\boldsymbol{\eta}_{1:d} \in A\right\} &= \mathbb{P}\left\{\boldsymbol{\beta}\left(\mathbf{U}_{1:d}\right) \in A\right\}, \end{aligned} \tag{D.52}$$

and $\beta_j\left(\mathbf{u}_{1:d}\right) \leq \alpha_j\left(\mathbf{r}_{1:d}\right)\left(\mathbf{u}_{1:d}\right)$ for every $\left(\mathbf{u}_{1:d}, j\right) \in (0,1)^d \times [d]$. The non-decreasing property of the set $A \subseteq \{\pm 1\}^d$ therefore implies

$$\left\{\mathbf{u}_{1:d} \in (0,1)^d : \boldsymbol{\beta}\left(\mathbf{u}_{1:d}\right) \in A\right\} \subseteq \left\{\mathbf{u}_{1:d} \in (0,1)^d : \boldsymbol{\alpha}\left(\mathbf{r}_{1:d}\right)\left(\mathbf{u}_{1:d}\right) \in A\right\} \tag{D.53}$$

Taking two observations equation D.52 and equation D.53 tells us that

$$\mathbb{P}\left\{\boldsymbol{\eta}_{1:d} \in A\right\} = \mathbb{P}\left\{\boldsymbol{\beta}\left(\mathbf{U}_{1:d}\right) \in A\right\} \leq \mathbb{P}\left\{\boldsymbol{\alpha}\left(\mathbf{r}_{1:d}\right)\left(\mathbf{U}_{1:d}\right) \in A\right\} = \mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in A \mid \mathbf{R}_{1:d} = \mathbf{r}_{1:d}\right\},$$

as claimed.

□

We are now in position to establish the first part equation D.3 of Lemma D.3. One can discover that

$$\begin{aligned} \mathbb{P}\left\{\sum_{j=1}^{d} T_j > 0\right\} &= \mathbb{P}\left\{\sum_{j=1}^{d} \varepsilon_j R_j > 0\right\} \\ &= \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in \mathcal{A}\left(\mathbf{R}_{1:d}\right) \mid \mathbf{R}_{1:d}\right\}\right] \\ &\overset{\text{(b)}}{\geq} \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{\eta}_{1:d} \in \mathcal{A}\left(\mathbf{R}_{1:d}\right) \mid \mathbf{R}_{1:d}\right\}\right] \\ &= \mathbb{E}\left[\mathbb{P}\left\{\sum_{j=1}^{d} \eta_j R_j > 0 \,\middle|\, \mathbf{R}_{1:d}\right\}\right] \\ &\overset{\text{(c)}}{=} \frac{1}{2}, \end{aligned} \tag{D.54}$$

where $\boldsymbol{\eta}_{1:d} \sim \text{Unif}\left(\{\pm 1\}^d\right) \perp\!\!\!\perp \mathbf{T}_{1:d}$, the step (b) follows due to Claim D.1, and the step (c) holds since while being conditioned on $\mathbf{R}_{1:d}$, we have

$$\sum_{j=1}^{d} \eta_j R_j \overset{d}{=} -\sum_{j=1}^{d} \eta_j R_j$$

and $\mathbb{P}\left\{\sum_{j=1}^{d} \eta_j R_j = 0\right\} = 0$. This concludes the proof of the first part equation D.3 of Lemma D.3.

Finally, it only remains to prove the second part equation D.4 of Lemma D.3. Towards this end, we define the event

$$\mathcal{G}_k := \left\{\mathbf{r}_{1:d} \in (0,+\infty)^d : r_k > \sum_{j \in [d] \setminus \{k\}} r_j\right\}.$$

If $\mathbf{r}_{1:d} \in \mathcal{G}_k$, then one can make the observation that

$$\{\varepsilon_k = +1\} \subseteq \left\{\sum_{j=1}^{d} \varepsilon_j r_j > 0\right\},$$

which therefore yields

$$\begin{aligned} \mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in \mathcal{A}\left(\mathbf{R}_{1:d}\right) \mid \mathbf{R}_{1:d} = \mathbf{r}_{1:d}\right\} \cdot \mathbb{1}_{\mathcal{G}_k}\left(\mathbf{r}_{1:d}\right) &\geq \mathbb{P}\left\{\varepsilon_k = +1 \middle| \mathbf{R}_{1:d} = \mathbf{r}_{1:d}\right\} \cdot \mathbb{1}_{\mathcal{G}_k}\left(\mathbf{r}_{1:d}\right) \\ &= p_k\left(r_k\right) \cdot \mathbb{1}_{\mathcal{G}_k}\left(\mathbf{r}_{1:d}\right). \end{aligned} \tag{D.55}$$

Therefore, we may conclude that

$$\begin{aligned}\mathbb{P}\left\{\sum_{j=1}^{d} T_j > 0\right\} &= \mathbb{P}\left\{\sum_{j=1}^{d} \varepsilon_j R_j > 0\right\} \\ &= \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in \mathcal{A}\left(\mathbf{R}_{1:d}\right) \mid \mathbf{R}_{1:d}\right\}\right] \\ &= \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in \mathcal{A}\left(\mathbf{R}_{1:d}\right) \mid \mathbf{R}_{1:d}\right\} \cdot \mathbb{1}_{\mathcal{G}_k}\left(\mathbf{R}_{1:d}\right)\right] \\ &\quad + \mathbb{E}\left[\mathbb{P}\left\{\boldsymbol{\varepsilon}_{1:d} \in \mathcal{A}\left(\mathbf{R}_{1:d}\right) \mid \mathbf{R}_{1:d}\right\} \cdot \mathbb{1}_{\mathcal{G}_k^c}\left(\mathbf{R}_{1:d}\right)\right] \\ &\geq \mathbb{E}\left[p_k\left(R_k\right) \cdot \mathbb{1}_{\mathcal{G}_k}\left(\mathbf{R}_{1:d}\right)\right] + \frac{1}{2}\mathbb{P}\left\{\mathbf{R}_{1:d} \in \mathcal{G}_k^c\right\} \\ &\overset{\text{(d)}}{>} \frac{1}{2}\mathbb{P}\left\{\mathbf{R}_{1:d} \in \mathcal{G}_k\right\} + \frac{1}{2}\mathbb{P}\left\{\mathbf{R}_{1:d} \in \mathcal{G}_k^c\right\} = \frac{1}{2}\end{aligned}$$

as desired, where the step (d) holds since $p_k\left(R_k\right) > \frac{1}{2}$ and $\mathbb{P}\left\{\mathbf{R}_{1:d} \in \mathcal{G}_k\right\} > 0$ from the assumptions.

E Omitted details and proofs for Section 5

E.1 Proof of Theorem 5.1

Before embarking on the proof of Theorem 5.1, let us introduce two sequences of random variables $\{\mathbb{L}_t\left(\mathbf{v}; n\right) : t \in \mathbb{Z}_+\}$ and $\{\mathbb{Q}_t\left(\mathbf{v}; n\right) : t \in \mathbb{Z}_+\}$, where

$$\mathbb{L}_t\left(\mathbf{v}; n\right) := \mathbf{v}^\top \widehat{\boldsymbol{\mu}}_t(n) \quad \text{and} \quad \mathbb{Q}_t\left(\mathbf{v}; n\right) := \mathbf{v}^\top \widehat{\boldsymbol{\Sigma}}_t(n)\mathbf{v} \tag{E.1}$$

for each fixed vector $\mathbf{v} \in \mathbb{R}^d$, which play a central role in the calculation of the characteristic function of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ for any fixed sample size $n \geq d+1$. In light of Proposition 3.1, one has for any $(T, \mathbf{v}) \in \mathbb{Z}_+ \times \left(\mathbb{R}^d \setminus \{\mathbf{0}_d\}\right)$ that

$$\mathbb{L}_{T+1}\left(\mathbf{v}; n\right) | \, \mathcal{F}_T(n) \sim \mathcal{N}\left(\mathbf{v}^\top \boldsymbol{\mu}_\star, \frac{1}{n}\sum_{t=0}^{T} \mathbb{Q}_t\left(\mathbf{v}; n\right)\right) \tag{E.2}$$

and

$$n \cdot \frac{\mathbb{Q}_{T+1}\left(\mathbf{v}; n\right)}{\mathbb{Q}_T\left(\mathbf{v}; n\right)} \Bigg| \, \mathcal{F}_T(n) \sim \chi^2(n-1). \tag{E.3}$$

Equipped with the above notation and properties, we are now ready to present our proof for Theorem 5.1.

Step 1: the characteristic function of the terminal mean vector. To begin with, let us consider the *covariance-path σ-field* defined on $\mathbb{R}^d \times \{\pm 1\}$:

$$\mathcal{V}(n) := \sigma\left(\left\{\widehat{\boldsymbol{\Sigma}}_t(n) : t \in \mathbb{Z}_+\right\}\right).$$

The first step is concerned with the conditional distribution of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$ given the covariance-path σ-field $\mathcal{V}(n)$ as a key basis for characterization of the unconditional distribution of $\widehat{\boldsymbol{\mu}}_\infty(n)$, which is accomplished through the following lemma.

Lemma E.1. *If $n \geq d+1$, then the following properties hold:*

(i) The matrix infinite sum $\sum_{T=0}^{\infty} \widehat{\boldsymbol{\Sigma}}_T(n)$ converges $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost surely to a $d \times d$ symmetric positive definite matrix.

(ii) The conditional distribution of the limiting point $\widehat{\boldsymbol{\mu}}_\infty(n)$ given the covariance-path σ-field $\mathcal{V}(n)$ is given by

$$\widehat{\boldsymbol{\mu}}_\infty(n) | \, \mathcal{V}(n) \sim \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{1}{n}\sum_{T=0}^{\infty} \widehat{\boldsymbol{\Sigma}}_T(n)\right). \tag{E.4}$$

Proof of Lemma E.1.

(i) It can be easily observed that

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\text{Trace}\left\{\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right\}\right] &= \sum_{T=0}^{\infty}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\text{Trace}\left\{\widehat{\boldsymbol{\Sigma}}_T(n)\right\}\right] \\ &\overset{\text{(a)}}{=} \sum_{T=0}^{\infty}\left(1-\frac{1}{n}\right)^T \text{Trace}\left(\boldsymbol{\Sigma}_\star\right) \\ &= n\cdot\text{Trace}\left(\boldsymbol{\Sigma}_\star\right) \\ &< +\infty,\end{aligned}$$

where the step (a) leverages the conditional contraction identity equation 3.3. Thus, it follows that

$$\text{Trace}\left\{\sum_{T=0}^{\infty}\hat{\boldsymbol{\Sigma}}_T(n)\right\} < +\infty \quad \mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\text{-almost surely,}$$

which in turn implies the $\mathbb{P}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}$-almost sure convergence of the matrix series $\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)$.

(ii) The consequence equation 3.2 of Proposition 3.1 tells us that

$$\widehat{\boldsymbol{\mu}}_T(n)|\,\mathcal{V}(n) \sim \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{1}{n}\sum_{t=0}^{T-1}\widehat{\boldsymbol{\Sigma}}_t(n)\right). \tag{E.5}$$

Thus, the conditional characteristic function of $\widehat{\boldsymbol{\mu}}_T(n)$ given the covariance-path σ-field $\mathcal{V}(n)$ equals to

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_T(n)\right\}\middle|\,\mathcal{V}(n)\right] = \exp\left[i\mathbf{v}^\top\boldsymbol{\mu}_\star - \frac{1}{2n}\mathbf{v}^\top\left\{\sum_{t=0}^{T-1}\widehat{\boldsymbol{\Sigma}}_t(n)\right\}\mathbf{v}\right]. \tag{E.6}$$

By virtue of Theorem 3.1, letting $T\to\infty$ to both sides of equation E.6 yields

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_\infty(n)\right\}\middle|\,\mathcal{V}(n)\right] &= \lim_{T\to\infty}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_T(n)\right\}\middle|\,\mathcal{V}(n)\right] \\ &= \lim_{T\to\infty}\exp\left[i\mathbf{v}^\top\boldsymbol{\mu}_\star - \frac{1}{2n}\mathbf{v}^\top\left\{\sum_{t=0}^{T-1}\widehat{\boldsymbol{\Sigma}}_t(n)\right\}\mathbf{v}\right] \\ &= \exp\left[i\mathbf{v}^\top\boldsymbol{\mu}_\star - \frac{1}{2n}\mathbf{v}^\top\left\{\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right\}\mathbf{v}\right],\end{aligned} \tag{E.7}$$

which concludes that

$$\widehat{\boldsymbol{\mu}}_\infty(n)|\,\mathcal{V}(n) \sim \mathcal{N}_d\left(\boldsymbol{\mu}_\star, \frac{1}{n}\sum_{T=0}^{\infty}\widehat{\boldsymbol{\Sigma}}_T(n)\right)$$

as desired.

□

In light of the conditional characteristic function of $\widehat{\boldsymbol{\mu}}_\infty(n)$ given the covariance-path σ-field $\mathcal{V}(n)$ from equation E.7, we define a function $\Phi_\infty^{(n)}:\mathbb{R}_+\to[0,1]$ as

$$\Phi_\infty^{(n)}(q) := \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{-\frac{1}{2n}\sum_{T=0}^{\infty}\mathbb{Q}_T\left(\mathbf{v};n\right)\right\}\middle|\,\mathbb{Q}_0\left(\mathbf{v};n\right)=q\right], \quad q\in\mathbb{R}_+, \tag{E.8}$$

so that the (unconditional) characteristic function of $\widehat{\boldsymbol{\mu}}_\infty(n)$ is written by

$$\text{ch}_{\widehat{\boldsymbol{\mu}}_\infty(n)}(\mathbf{v}) := \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left\{i\mathbf{v}^\top\widehat{\boldsymbol{\mu}}_\infty(n)\right\}\right] = \exp\left(i\mathbf{v}^\top\boldsymbol{\mu}_\star\right)\Phi_\infty^{(n)}\left(\mathbf{v}^\top\boldsymbol{\Sigma}_\star\mathbf{v}\right) \tag{E.9}$$

for $\mathbf{v} \in \mathbb{R}^d$.

We next step forward to take a closer look at the real part $\Phi_\infty^{(n)} : \mathbb{R}_+ \to [0,1]$ of the characteristic function of the converging point $\widehat{\boldsymbol{\mu}}_\infty(n)$. Towards this end, we fix any $\mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ and then define $Z_t(\mathbf{v}; n) := n \cdot \frac{\mathbb{Q}_{t+1}(\mathbf{v};n)}{\mathbb{Q}_t(\mathbf{v};n)}$ for $t \in \mathbb{Z}_+$. By the property equation E.3, we have

$$\{Z_t(\mathbf{v}; n) : t \in \mathbb{Z}_+\} \overset{\text{i.i.d.}}{\sim} \chi^2(n-1) \overset{d}{=} \Gamma\left(\frac{n-1}{2}, 2\right). \tag{E.10}$$

On the other hand, the definition of the sequence $\{Z_t(\mathbf{v}; n) : t \in \mathbb{Z}_+\}$ yields for every $T \in \mathbb{N}$ that

$$\mathbb{Q}_T(\mathbf{v}; n) = \mathbb{Q}_0(\mathbf{v}; n) \cdot \prod_{t=0}^{T-1} \frac{Z_t(\mathbf{v}; n)}{n} = \mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v} \cdot \prod_{t=0}^{T-1} \frac{Z_t(\mathbf{v}; n)}{n}. \tag{E.11}$$

We now introduce the *random perpetuity*:

$$\mathfrak{R}(\mathbf{v}; n) := 1 + \sum_{T=1}^{\infty} \left\{ \prod_{t=0}^{T-1} \frac{Z_t(\mathbf{v}; n)}{n} \right\} \in [1, +\infty], \tag{E.12}$$

which turns out to be finite $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost surely, since

$$\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}[\mathfrak{R}(\mathbf{v}; n)] = \frac{1}{\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}} \sum_{T=0}^{\infty} \underbrace{\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}[\mathbb{Q}_T(\mathbf{v}; n)]}_{= \mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v} \left(1 - \frac{1}{n}\right)^T} = n < +\infty.$$

With the notion of the random perpetuity equation E.12 in place, one can write the real part of the characteristic function of $\widehat{\boldsymbol{\mu}}_\infty(n)$ from the equation equation E.9 as

$$\Phi_\infty^{(n)}\left(\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}\right) = \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\exp\left\{-\frac{\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}{2} \cdot \frac{\mathfrak{R}(\mathbf{v}; n)}{n}\right\}\right], \quad \mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}. \tag{E.13}$$

Step 2: weak convergence of $\frac{\mathfrak{R}(\mathbf{v};n)}{n}$ as $n \to \infty$. Next, we isolate the following result to guarantee the weak convergence of $\frac{\mathfrak{R}(\mathbf{v};n)}{n}$ as $n \to \infty$ whose weak limit is independent of $\mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$. This is because the random perpetuity $\mathfrak{R}(\mathbf{v}; n)$ is a function of the sequence $(Z_t(\mathbf{v}; n) : t \in \mathbb{Z}_+)$ of i.i.d. random variables whose distribution is independent of $\mathbf{v}$. We defer its proof to Appendix E.2.

Theorem E.1. *It holds for any $\mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ that*

$$\frac{\mathfrak{R}(\mathbf{v}; n)}{n} \Rightarrow \mathfrak{I} := \int_0^\infty \exp\left\{\sqrt{2}\mathbb{B}(t) - 2t\right\} \mathrm{d}t \tag{E.14}$$

as $n \to \infty$, where $(\mathbb{B}(t) : t \in \mathbb{R}_+)$ is a standard Brownian motion (the Wiener process) in $\mathbb{R}$.

Step 3: distributional characterization for the weak limit $\mathfrak{I}$. The following theorem characterizes the exact distribution of the weak limit of $\frac{\mathfrak{R}(\mathbf{v};n)}{n}$ as $n \to \infty$ introduced in Theorem E.1.

Theorem E.2. *We have*

$$\mathfrak{I} = \int_0^\infty \exp\left\{\sqrt{2}\mathbb{B}(t) - 2t\right\} \mathrm{d}t \overset{d}{=} \frac{1}{G}, \quad G \sim \Gamma(2, 1). \tag{E.15}$$

The proof of Theorem E.2 will be elaborated in Appendix E.3.

Step 4: putting everything together. Now, we are in position to finish the proof of Theorem 5.1. By letting $n \to \infty$ in the equation equation E.13, Theorems E.1 and E.2 reveal that

$$\begin{aligned}
\lim_{n\to\infty} \Phi_\infty^{(n)}\left(\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}\right) &= \lim_{n\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\exp\left\{-\frac{\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}{2} \cdot \frac{\mathfrak{R}(\mathbf{v}; n)}{n}\right\}\right] \\
&\overset{\text{(a)}}{=} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[\exp\left(-\frac{\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}{2} \cdot \mathfrak{I}\right)\right] \\
&= \mathbb{E}_{G \sim \Gamma(2,1)}\left[\exp\left(-\frac{\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}{2G}\right)\right] \\
&\overset{\text{(b)}}{=} \mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v} \cdot K_2\left(\sqrt{2\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}\right)
\end{aligned} \tag{E.16}$$

for any fixed $\mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$, where the step (a) follows due to the continuous mapping theorem, and the step (b) makes use of the equation equation E.102 with $q = \frac{\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}{2}$. Substitution of the equation equation E.16 into equation E.9 then yields

$$\begin{aligned} \lim_{n\to\infty} \mathrm{ch}_{\widehat{\boldsymbol{\mu}}_\infty(n)}(\mathbf{v}) &= \exp\left(i\mathbf{v}^\top \boldsymbol{\mu}_\star\right) \cdot \lim_{n\to\infty} \Phi_\infty^{(n)}\left(\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}\right) \\ &= \exp\left(i\mathbf{v}^\top \boldsymbol{\mu}_\star\right) \cdot \mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v} \cdot K_2\left(\sqrt{2\mathbf{v}^\top \boldsymbol{\Sigma}_\star \mathbf{v}}\right) \end{aligned} \tag{E.17}$$

for every $\mathbf{v} \in \mathbb{R}^d$. On the other hand, it is well-known that the characteristic function of $t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ equals to

$$\mathrm{ch}_{t_d(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}(\mathbf{v}) = \exp\left(i\mathbf{v}^\top \boldsymbol{\mu}\right) \frac{2^{1-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \left(\sqrt{\nu \cdot \mathbf{v}^\top \boldsymbol{\Sigma} \mathbf{v}}\right)^{\frac{\nu}{2}} K_2\left(\sqrt{\nu \cdot \mathbf{v}^\top \boldsymbol{\Sigma} \mathbf{v}}\right), \quad \mathbf{v} \in \mathbb{R}^d. \tag{E.18}$$

By taking $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \left(\boldsymbol{\mu}_\star, \frac{1}{2}\boldsymbol{\Sigma}_\star, 4\right)$ and comparing two equations equation E.17 and equation E.18, one can reveal that

$$\lim_{n\to\infty} \mathrm{ch}_{\widehat{\boldsymbol{\mu}}_\infty(n)}(\mathbf{v}) = \mathrm{ch}_{t_d\left(\boldsymbol{\mu}_\star, \frac{1}{2}\boldsymbol{\Sigma}_\star, 4\right)}(\mathbf{v}), \quad \mathbf{v} \in \mathbb{R}^d,$$

which concludes the proof of $\widehat{\boldsymbol{\mu}}_\infty(n) \Rightarrow t_d\left(\boldsymbol{\mu}_\star, \frac{1}{2}\boldsymbol{\Sigma}_\star, 4\right)$ as $n \to \infty$.

E.2 Proof of Theorem E.1

Before proceeding, let us introduce $X_t(\mathbf{v}; n) := \log Z_t(\mathbf{v}; n) - \log n$ for $t \in \mathbb{Z}_+$ and the *log-random walk* $\{S_t(\mathbf{v}; n) : t \in \mathbb{Z}_+\}$ defined as

$$S_0(\mathbf{v}; n) := 0 \quad \text{and} \quad S_t(\mathbf{v}; n) := \sum_{j=0}^{t-1} X_j(\mathbf{v}; n), \quad t \in \mathbb{N}, \tag{E.19}$$

which then yields

$$\frac{\mathfrak{R}(\mathbf{v}; n)}{n} = \frac{1}{n} \sum_{t=0}^{\infty} \exp\left\{S_t(\mathbf{v}; n)\right\}. \tag{E.20}$$

The following lemma provides the exact formulae of the mean and variance of $X_t(\mathbf{v}; n)$ and their asymptotic expansions as $n \to \infty$.

Lemma E.2 (Exact formulae and asymptotic expansions of the mean and variance of $X_t(\mathbf{v}; n)$)**.** *It holds that*

$$\mu_n := \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[X_0(\mathbf{v}; n)\right] = \psi\left(\frac{n-1}{2}\right) + \log 2 - \log n, \tag{E.21}$$

and

$$\sigma_n^2 := \mathrm{Var}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[X_0(\mathbf{v}; n)\right] = \psi_1\left(\frac{n-1}{2}\right), \tag{E.22}$$

where the digamma function $\psi : (0, +\infty) \to \mathbb{R}$ and the trigamma function $\psi_1 : (0, +\infty) \to \mathbb{R}$ are defined as

$$\psi(x) := \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)} \quad \text{and} \quad \psi_1(x) := \psi'(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \log \Gamma(x),$$

respectively. Here, the gamma function $\Gamma : (0, +\infty) \to \mathbb{R}$ is defined to be

$$\Gamma(x) := \int_0^\infty t^{x-1} \exp(-t)\, \mathrm{d}t, \quad x \in (0, +\infty).$$

Also, we have

$$\mu_n = -\frac{2}{n} + \mathcal{O}\left(\frac{1}{n^2}\right) \quad \text{and} \quad \sigma_n^2 = \frac{2}{n} + \mathcal{O}\left(\frac{1}{n^2}\right) \tag{E.23}$$

as $n \to \infty$.

Proof of Lemma E.2. It is widely known that if $Z \sim \Gamma(\alpha, \theta)$, then

$$\mathbb{E}[\log Z] = \psi(\alpha) + \log\theta \quad \text{and} \quad \text{Var}[\log Z] = \psi_1(\alpha). \tag{E.24}$$

The exact formulae equation E.21 and equation E.22 for the mean and variance of $X_t(\mathbf{v}; n)$, respectively, directly follows by putting $(\alpha, \theta) = \left(\frac{n-1}{2}, 2\right)$ into the equation equation E.24.

Let us next turn to the asymptotic expansions of μ_n and σ_n^2 as $n \to \infty$. To this end, one can reveal the following classical asymptotic expansions of the digamma function and trigamma function: as $\alpha \to \infty$, one has

$$\psi(\alpha) = \log\alpha - \frac{1}{2\alpha} + \mathcal{O}\left(\frac{1}{\alpha^2}\right) \quad \text{and} \quad \psi_1(\alpha) = \frac{1}{\alpha} + \mathcal{O}\left(\frac{1}{\alpha^2}\right). \tag{E.25}$$

Letting $\alpha = \alpha_n := \frac{n-1}{2}$ in the asymptotic expansions of equation E.25, it follows that as $n \to \infty$,

$$\psi(\alpha_n) = \log\left(\frac{n-1}{2}\right) - \frac{1}{n-1} + \mathcal{O}\left(n^{-2}\right) \quad \text{and} \quad \psi_1(\alpha_n) = \frac{2}{n} + \mathcal{O}\left(n^{-2}\right). \tag{E.26}$$

By taking the asymptotic expansions of equation E.26 collectively into the exact formulae equation E.21 and equation E.22 collectively, we reach

$$\mu_n = \psi(\alpha_n) + \log 2 - \log n = \log\left(1 - \frac{1}{n}\right) - \frac{1}{n-1} + \mathcal{O}\left(n^{-2}\right) = -\frac{2}{n} + \mathcal{O}\left(n^{-2}\right)$$

and

$$\sigma_n^2 = \psi_1(\alpha_n) = \frac{2}{n} + \mathcal{O}\left(n^{-2}\right)$$

as $n \to \infty$, which completes the proof of Lemma E.2.

□

With the above preliminaries in place, we next fix any $\mathbf{v} \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}$ and define a continuous-time càdlàg process $W(\mathbf{v}; n) : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ for $n > 1$ by

$$W(\mathbf{v}; n)(\omega, t) := S_{\lfloor nt \rfloor}(\mathbf{v}; n)(\omega), \tag{E.27}$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the underlying probability space. Also, define a continuous-time stochastic process $W : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ as

$$W(\omega, t) := \sqrt{2}\mathbb{B}(\omega, t) - 2t. \tag{E.28}$$

By the fact that $\mathbb{P}\{\omega \in \Omega : W(\omega, \cdot) \in \mathbb{D}([0, +\infty))\} = 1$, it can also be regarded as a continuous-time càdlàg process $\mathbb{P}$-almost surely. By viewing the continuous-time càdlàg processes $W(\mathbf{v}; n) : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ and $W : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ as the $\mathbb{D}([0, +\infty))$-valued random elements $W(\mathbf{v}; n) : \Omega \to \mathbb{D}([0, +\infty))$ and $W : \Omega \to \mathbb{D}([0, +\infty))$, we aim to prove the following functional CLT (invariance principle) for the sequence $\{W(\mathbf{v}; n) \in (\Omega \to \mathbb{D}([0, +\infty))) : n \in \mathbb{N} \setminus \{1\}\}$.

Lemma E.3 (Functional CLT for $W(\mathbf{v}; n)$)**.** *The following weak convergence in* $\mathbb{D}([0, +\infty))$ *holds as* $n \to \infty$*:*

$$W(\mathbf{v}; n) \Rightarrow W. \tag{E.29}$$

Proof of Lemma E.3. By appealing to Lemma A.1, it is sufficient to establish the weak convergence $\pi_T \circ W(\mathbf{v}; n) \Rightarrow \pi_T \circ W$ in $\mathbb{D}([0, T])$ as $n \to \infty$ for an arbitrary $T \in (0, +\infty)$. Towards achieving this goal, we decompose the continuous-time càdlàg process $W(\mathbf{v}; n)$ into a deterministic drift part plus a martingale. In order to do so, define a discrete-time filtration $(\mathcal{G}_k(\mathbf{v}; n) : k \in \mathbb{Z}_+)$ on $(\Omega, \mathcal{F})$ as $\mathcal{G}_0(\mathbf{v}; n) := \{\varnothing, \Omega\}$ and

$$\mathcal{G}_k(\mathbf{v}; n) := \sigma\left(\{Z_0(\mathbf{v}; n), Z_1(\mathbf{v}; n), \cdots, Z_{k-1}(\mathbf{v}; n)\}\right), \quad k \in \mathbb{N},$$

together with the centered increments $\Delta M_k(\mathbf{v};n) := X_k(\mathbf{v};n) - \mu_n$ for $k \in \mathbb{Z}_+$. With these notions in hand, we now define the càdlàg step martingale $M(\mathbf{v};n) : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ and the deterministic drift term $A_n \in \mathbb{D}([0,+\infty))$ by

$$M(\mathbf{v};n)(\omega,t) := \sum_{k=0}^{\lfloor nt \rfloor - 1} \Delta M_k(\mathbf{v};n)(\omega) \quad \text{and} \quad A_n(t) := \lfloor nt \rfloor \mu_n. \tag{E.30}$$

Note that $M(\mathbf{v};n) : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is a $\mathcal{G}_n(\star)$-martingale, where

$$\mathcal{G}_n(\star) := \left(\mathcal{G}_n(t) := \mathcal{G}_{\lfloor nt \rfloor}(n) : t \in \mathbb{R}_+\right), \quad n \in \mathbb{N} \setminus \{1\},$$

and taken together with the drift term $A_n : \mathbb{R}_+ \to \mathbb{R}$, it satisfies the drift-martingale decomposition of $W(\mathbf{v};n)$:

$$W(\mathbf{v};n)(\omega,t) = M(\mathbf{v};n)(\omega,t) + A_n(t), \quad (\omega,t) \in \Omega \times \mathbb{R}_+. \tag{E.31}$$

Step 1: analysis of the martingale part $M(\mathbf{v};n)$. Here, we aim to prove for any $T \in (0,+\infty)$ that in $\mathbb{D}([0,T])$,

$$\pi_T \circ M(\mathbf{v};n) \Rightarrow \sqrt{2}\,(\pi_T \circ \mathbb{B}) \tag{E.32}$$

as $n \to \infty$, with the aid of the continuous-time martingale functional CLT (Lemma A.9). Towards this end, we only need to verify that the two sufficient conditions of Lemma A.9 hold for the sequence $\{M(\mathbf{v};n) : n \in \mathbb{N} \setminus \{1\}\}$ of càdlàg, pure-jump, and square-integrable $\mathcal{G}_n(\star)$-martingales. We begin with the derivation of the $\mathcal{G}_n(\star)$-predictable quadratic variation $\langle M(\mathbf{v};n)\rangle : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ for the pure-jump and square-integrable $\mathcal{G}_n(\star)$-martingale $M(\mathbf{v};n) : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ using Lemma A.3:

$$\begin{aligned}
\langle M(\mathbf{v};n)\rangle(t) &= \sum_{s \in (0,t]} \mathbb{E}\left[\{\Delta M(\mathbf{v};n)(s)\}^2 \middle| \mathcal{G}_n(s-)\right] \\
&\overset{\text{(a)}}{=} \sum_{j=1}^{\lfloor nt \rfloor} \mathbb{E}\left[\left\{\Delta M(\mathbf{v};n)\left(\frac{j}{n}\right)\right\}^2 \middle| \mathcal{G}_n\left(\frac{j}{n}-\right)\right] \\
&\overset{\text{(b)}}{=} \sum_{j=0}^{\lfloor nt \rfloor - 1} \mathbb{E}\left[\{\Delta M_j(\mathbf{v};n)\}^2 \middle| \mathcal{G}_j(n)\right] \\
&\overset{\text{(c)}}{=} \sum_{j=0}^{\lfloor nt \rfloor - 1} \mathbb{E}\left[\{\Delta M_j(\mathbf{v};n)\}^2\right] \\
&= \lfloor nt \rfloor \sigma_n^2
\end{aligned} \tag{E.33}$$

$\mathbb{P}$-almost surely, where the step (a) holds due to the fact that

$$\{s \in (0,t] : \Delta M(\mathbf{v};n)(s) \neq 0\} = \left\{\frac{j}{n} : j \in [\lfloor nt \rfloor]\right\}, \tag{E.34}$$

the step (b) follows since $\Delta M(\mathbf{v};n)\left(\frac{j}{n}\right) = \Delta M_{j-1}(\mathbf{v};n)$ and $\mathcal{G}_n\left(\frac{j}{n}-\right) = \mathcal{G}_{j-1}(n)$ for $j \in [\lfloor nt \rfloor]$, and the step (c) leverages the observation that $\{\Delta M_k(\mathbf{v};n) : k \in \mathbb{Z}_+\}$ are i.i.d. random variables.

(i) **Predictable quadratic variation convergence**: Leveraging the derivation equation E.33, one can reveal that

$$\begin{aligned}
&\sup\{|\langle M(\mathbf{v};n)\rangle(t) - 2t| : t \in [0,T]\} \\
&\overset{\text{(d)}}{=} \sup\left\{\left|\lfloor nt \rfloor \sigma_n^2 - 2t\right| : t \in [0,T]\right\} \\
&\overset{\text{(e)}}{=} \sup\left\{\left|\lfloor nt \rfloor \left\{\frac{2}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right\} - 2t\right| : t \in [0,T]\right\} \\
&\leq \sup\left\{\frac{2}{n}(nt - \lfloor nt \rfloor) + \lfloor nt \rfloor \cdot \mathcal{O}\left(\frac{1}{n^2}\right) : t \in [0,T]\right\} \\
&\leq \frac{2}{n} + T \cdot \mathcal{O}\left(\frac{1}{n}\right) \overset{n \to \infty}{\longrightarrow} 0,
\end{aligned}$$

where the step (d) holds $\mathbb{P}$-almost surely, and the step (e) follows by using the asymptotic expansion equation E.23 of σ_n^2 from Lemma E.2. Hence, it follows that

$$\sup\left\{\left|\langle M(\mathbf{v};n)\rangle(t)-2t\right| : t\in[0,T]\right\} \overset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}\text{-almost surely.} \tag{E.35}$$

(ii) **Lindeberg condition**: Given any $\varepsilon\in(0,+\infty)$, let

$$\xi_n(\varepsilon;\mathbf{v}) := \sum_{t\in(0,T]} \mathbb{E}\left[\left\{\Delta M(\mathbf{v};n)(t)\right\}^2 \cdot \mathbb{1}_{\{|\Delta M(\mathbf{v};n)(t)|>\varepsilon\}} \middle| \mathcal{G}_n(t-)\right].$$

Then, it can be observed that

$$\begin{aligned}
\xi_n(\varepsilon;\mathbf{v}) &\overset{\text{(f)}}{=} \sum_{j=1}^{\lfloor nT\rfloor} \mathbb{E}\left[\left\{\Delta M(\mathbf{v};n)\left(\frac{j}{n}\right)\right\}^2 \cdot \mathbb{1}_{\left\{\left|\Delta M(\mathbf{v};n)\left(\frac{j}{n}\right)\right|>\varepsilon\right\}} \middle| \mathcal{G}_n\left(\frac{j}{n}-\right)\right] \\
&\overset{\text{(g)}}{=} \sum_{j=0}^{\lfloor nT\rfloor-1} \mathbb{E}\left[\left\{\Delta M_k(\mathbf{v};n)\right\}^2 \cdot \mathbb{1}_{\{|\Delta M_k(\mathbf{v};n)|>\varepsilon\}} \middle| \mathcal{G}_j(n)\right] \\
&\overset{\text{(h)}}{=} \lfloor nT\rfloor\, \mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^2 \cdot \mathbb{1}_{\{|\Delta M_0(\mathbf{v};n)|>\varepsilon\}}\right],
\end{aligned} \tag{E.36}$$

where the step (f) comes from equation E.34, the step (g) utilizes the fact that $\Delta M(\mathbf{v};n)\left(\frac{j}{n}\right) = \Delta M_{j-1}(\mathbf{v};n)$ and $\mathcal{G}_n\left(\frac{j}{n}-\right) = \mathcal{G}_{j-1}(n)$ for each $j\in[\lfloor nt\rfloor]$, and the step (h) follows since $\{\Delta M_k(\mathbf{v};n) : k\in\mathbb{Z}_+\}$ are i.i.d. random variables. We now aim to establish that

$$\lim_{n\to\infty} n\cdot\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^2 \cdot \mathbb{1}_{\{|\Delta M_0(\mathbf{v};n)|>\varepsilon\}}\right] = 0, \tag{E.37}$$

which immediately verifies the Lindeberg condition $\xi_n(\varepsilon;\mathbf{v}) \xrightarrow{p} 0$ as $n\to\infty$ by virtue of the identity equation E.36. By the Cauchy-Schwarz inequality and Markov's inequality, we have

$$\begin{aligned}
&n\cdot\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^2 \cdot \mathbb{1}_{\{|\Delta M_0(\mathbf{v};n)|>\varepsilon\}}\right] \\
&\le n\left(\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^4\right]\right)^{\frac{1}{2}} \underbrace{\left(\mathbb{P}\left\{|\Delta M_0(\mathbf{v};n)|>\varepsilon\right\}\right)^{\frac{1}{2}}}_{\le\, \varepsilon^{-2}\left(\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^4\right]\right)^{\frac{1}{2}}} \\
&\le \frac{n}{\varepsilon^2}\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^4\right].
\end{aligned} \tag{E.38}$$

In light of the inequality equation E.38, let us claim that $\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^4\right] = \mathcal{O}\left(n^{-2}\right)$. Towards proving this claim, we first observe that

$$\begin{aligned}
\mathbb{E}\left[\left\{\Delta M_0(\mathbf{v};n)\right\}^4\right] &= \mathbb{E}\left[\left\{X_0(\mathbf{v};n)-\mu_n\right\}^4\right] \\
&\overset{\text{(i)}}{\le} 8\mathbb{E}\left[\left\{X_0(\mathbf{v};n)\right\}^4\right] + 8\mu_n^4 \\
&\overset{\text{(j)}}{=} 8\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\log^4\left(\frac{Z}{n}\right)\right] + \mathcal{O}\left(n^{-4}\right),
\end{aligned} \tag{E.39}$$

where the step (i) results from the simple inequality $(a+b)^4 \le 8\left(a^4+b^4\right)$ for $(a,b)\in\mathbb{R}\times\mathbb{R}$, and the step (j) is due to the fact $|\mu_n| = \mathcal{O}\left(n^{-1}\right)$, which follows from Lemma E.2. We now turn our attention to bound the term $\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\log^4\left(\frac{Z}{n}\right)\right]$. Define the *relative fluctuation*

$$U := \frac{Z-(n-1)}{n-1}, \quad Z\sim\chi^2(n-1).$$

One can easily reveal that the moments of the relative fluctuation U are given by

$$\begin{aligned}
\mathbb{E}\left[U^2\right] &= \frac{2}{n-1} = \mathcal{O}\left(n^{-1}\right), \\
\mathbb{E}\left[U^4\right] &= \frac{12}{(n-1)^2} + \frac{48}{(n-1)^3} = \mathcal{O}\left(n^{-2}\right), \\
\mathbb{E}\left[U^6\right] &= \frac{120}{(n-1)^3} + \frac{2080}{(n-1)^4} + \frac{3840}{(n-1)^5} = \mathcal{O}\left(n^{-2}\right), \\
\mathbb{E}\left[U^8\right] &= \frac{1680}{(n-1)^4} + \frac{76160}{(n-1)^5} + \frac{295680}{(n-1)^6} + \frac{645120}{(n-1)^7} = \mathcal{O}\left(n^{-4}\right)
\end{aligned} \tag{E.40}$$

Utilizing the identity $\log\left(\frac{Z}{n}\right) = \log\left(1-\frac{1}{n}\right) + \log(1+U)$ together with the simple inequality $(a+b)^4 \le 8\left(a^4+b^4\right)$ for $(a,b) \in \mathbb{R}\times\mathbb{R}$, we obtain for any $n \ge 2$ that

$$\begin{aligned}
\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\log^4\left(\frac{Z}{n}\right)\right] &\le 8\left|\log\left(1-\frac{1}{n}\right)\right|^4 + 8\mathbb{E}\left[\log^4(1+U)\right] \\
&\overset{\text{(k)}}{\le} 128n^{-4} + 8\mathbb{E}\left[\log^4(1+U)\right],
\end{aligned} \tag{E.41}$$

where the step (k) follows since $\left|\log\left(1-\frac{1}{n}\right)\right| = \log\left(1+\frac{1}{n-1}\right) \le \frac{1}{n-1} \le \frac{2}{n}$ for all $n \ge 2$. Thus, we only need to control $\mathbb{E}\left[\log^4(1+U)\right]$ to upper bound $\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\log^4\left(\frac{Z}{n}\right)\right]$. Let us first control $|\log(1+U)|$ on the event $\left\{|U| \le \frac{1}{2}\right\}$:

$$\begin{aligned}
&\mathbb{E}\left[\log^4(1+U)\right] \\
&= \mathbb{E}\left[|\log(1+U)|^4 \cdot \mathbb{1}_{\left\{|U|\le\frac{1}{2}\right\}}\right] + \mathbb{E}\left[|\log(1+U)|^4 \cdot \mathbb{1}_{\left\{|U|>\frac{1}{2}\right\}}\right] \\
&\overset{\text{(l)}}{\le} (2\log 2)^4\,\mathbb{E}\left[U^4\right] + \left(\mathbb{E}\left[|\log(1+U)|^8\right]\right)^{\frac{1}{2}}\left(\mathbb{P}\left\{|U|>\frac{1}{2}\right\}\right)^{\frac{1}{2}} \\
&\overset{\text{(m)}}{\le} (2\log 2)^4\,\mathbb{E}\left[U^4\right] + \left(\mathbb{E}\left[|\log(1+U)|^8\right]\right)^{\frac{1}{2}} \cdot 2^4\left(\mathbb{E}\left[U^8\right]\right)^{\frac{1}{2}}
\end{aligned} \tag{E.42}$$

where the step (l) follows by the simple inequality $|\log(1+x)| \le 2\log 2|x|$ for $|x| \le \frac{1}{2}$ together with the Cauchy-Schwarz inequality, and the step (m) uses Markov's inequality.

Claim E.1. *As $n \to \infty$, we have*

$$\mathbb{E}\left[|\log(1+U)|^8\right] = \mathbb{E}_{Z\sim\chi^2(n-1)}\left[\left|\log\left(\frac{Z}{n-1}\right)\right|^8\right] = \mathcal{O}(1). \tag{E.43}$$

Proof of Claim E.1. The proof of Claim E.1 heavily relies upon the following basic facts: there exists an absolute constant $c \in (0,+\infty)$ such that

$$|\log x|^8 \le c\left(x^{\frac{1}{4}} + x^{-\frac{1}{4}}\right), \quad x \in (0,+\infty),$$

which immediately gives

$$\begin{aligned}
&\mathbb{E}\left[|\log(1+U)|^8\right] \\
&\le c\left\{\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\left(\frac{Z}{n-1}\right)^{\frac{1}{4}}\right] + \mathbb{E}_{Z\sim\chi^2(n-1)}\left[\left(\frac{Z}{n-1}\right)^{-\frac{1}{4}}\right]\right\}
\end{aligned} \tag{E.44}$$

and $\chi^2(k) \overset{d}{=} \Gamma\left(\frac{k}{2}, 2\right)$ and thus it follows that

$$\mathbb{E}_{Z\sim\chi^2(k)}\left[Z^r\right] = 2^r \cdot \frac{\Gamma\left(\frac{k}{2}+r\right)}{\Gamma\left(\frac{k}{2}\right)}, \quad r \in \left(-\frac{k}{2}, +\infty\right). \tag{E.45}$$

By taking $(k, r) = \left(n-1, \frac{1}{4}\right)$ in the identity equation E.45, we have

$$\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\left(\frac{Z}{n-1}\right)^{\frac{1}{4}}\right] = \left(\frac{n-1}{2}\right)^{-\frac{1}{4}} \cdot \frac{\Gamma\left(\frac{n-1}{2}+\frac{1}{4}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \overset{\text{(n)}}{=} \mathcal{O}(1) \tag{E.46}$$

as $n \to \infty$, where the step (n) follows due to the following well-known asymptotic relation:

$$\lim_{x\to+\infty} \frac{\Gamma(x+\alpha)}{x^{\alpha}\Gamma(x)} = 1, \quad \alpha \in \mathbb{R}. \tag{E.47}$$

On the other hand, putting $(k, r) = \left(n-1, -\frac{1}{4}\right)$ into the identity equation E.45 yields

$$\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\left(\frac{Z}{n-1}\right)^{-\frac{1}{4}}\right] = \left(\frac{n-1}{2}\right)^{\frac{1}{4}} \cdot \frac{\Gamma\left(\frac{n-1}{2}-\frac{1}{4}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \overset{\text{(o)}}{=} \mathcal{O}(1) \tag{E.48}$$

as $n \to \infty$, where the step (o) is a consequence of the key asymptotic relation equation E.47. Hence, taking two pieces equation E.46 and equation E.48 collectively into the bound equation E.44 completes the proof of Claim E.1.

□

Appealing to Claim E.1 together with the facts $\mathbb{E}\left[U^4\right] = \mathcal{O}\left(n^{-2}\right)$ and $\mathbb{E}\left[U^8\right] = \mathcal{O}\left(n^{-4}\right)$ from equation E.40, the bound equation E.42 tells us that

$$\mathbb{E}\left[\log^4(1+U)\right] = \mathcal{O}\left(n^{-2}\right) \tag{E.49}$$

as $n \to \infty$. Combining the bound equation E.49 together with the inequality equation E.41, one also has

$$\mathbb{E}_{Z\sim\chi^2(n-1)}\left[\log^4\left(\frac{Z}{n}\right)\right] = \mathcal{O}\left(n^{-2}\right) \tag{E.50}$$

Next, taking equation E.50 into the initial-step bound equation E.39 finishes the proof of the desired claim

$$\mathbb{E}\left[\{\Delta M_0(\mathbf{v}; n)\}^4\right] = \mathcal{O}\left(n^{-2}\right) \tag{E.51}$$

as $n \to \infty$. Hence, taking two bounds equation E.38 and equation E.51 collectively deduces the original claim equation E.37, which completes the verification of the Lindeberg condition $\xi_n(\varepsilon; \mathbf{v}) \xrightarrow{p} 0$ as $n \to \infty$.

Step 2: analysis of the deterministic drift term A_n. Invoking the asymptotic expansion of μ_n from Lemma E.2, we obtain

$$A_n(t) + 2t = \lfloor nt \rfloor \mu_n + 2t = \frac{2}{n}(nt - \lfloor nt \rfloor) + \lfloor nt \rfloor \cdot \mathcal{O}\left(\frac{1}{n^2}\right).$$

Thus, it follows that

$$\sup\{|A_n(t) + 2t| : t \in [0, T]\} \leq \frac{2}{n} + T \cdot \mathcal{O}\left(\frac{1}{n}\right) \to 0 \tag{E.52}$$

as $n \to \infty$. Hence, we arrive from the fact equation E.52 that in the Skorokhod space $\mathbb{D}([0, T])$, we have

$$\pi_T(A_n) \overset{n\to\infty}{\longrightarrow} -2\pi_T(I_{\mathbb{R}}), \tag{E.53}$$

where $I_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ is the identity function on the real line $\mathbb{R}$.

Step 3: putting everything together. Finally, it is time to put everything together. Due to Slutsky's theorem for the Skorokhod space (Lemma A.2), one can find from equation E.32 and equation E.53 that for every $T \in (0, +\infty)$,

$$\pi_T \circ W(\mathbf{v}; n) = \pi_T \circ \{M(\mathbf{v}; n) + A_n\} \Rightarrow \pi_T \circ \left(\sqrt{2}\mathbb{B} - 2I_{\mathbb{R}}\right) = \pi_T \circ W \tag{E.54}$$

in the Skorokhod space $\mathbb{D}([0, T])$, as $n \to \infty$. This concludes the proof of Lemma E.3.

□

With the functional CLT for the sequence $\{W(\mathbf{v}; n) : n \in \mathbb{N} \setminus \{1\}\}$ (Lemma E.3) in place, we are now in position to establish the following finite-horizon convergence guarantee.

Lemma E.4 (Finite-horizon convergence)**.** *Given any fixed $M \in (0, +\infty)$, we define*

$$\mathcal{C}_n(\mathbf{v}; M) := \frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\{S_t(\mathbf{v}; n)\} \tag{E.55}$$

and

$$\mathcal{C}(M) := \int_0^M \exp\{W(t)\}\, \mathrm{d}t, \tag{E.56}$$

where $W : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is the continuous-time ($\mathbb{P}$-almost surely) càdlàg stochastic process defined in equation E.28. Then,

$$\mathcal{C}_n(\mathbf{v}; M) \Rightarrow \mathcal{C}(M) \tag{E.57}$$

as $n \to \infty$.

Proof of Lemma E.4. We aim to prove Lemma E.4 by leveraging the extended continuous mapping theorem (Lemma A.8). In order to do so, let us first define the functionals $F_n : \mathbb{D}([0, M]) \to \mathbb{R}$ for $n \in \mathbb{N}$ and $F : \mathbb{D}([0, M]) \to \mathbb{R}$ by

$$F_n(x) := \frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x\left(\frac{t}{n}\right)\right\} \quad \text{and} \quad F(x) := \int_0^M \exp\{x(t)\}\, \mathrm{d}t, \tag{E.58}$$

respectively. We set $(\mathbb{D}, d) = (\mathbb{D}([0, M]), \sigma)$ and $(\mathbb{E}, e) = (\mathbb{R}, d_1)$, where $d_1 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ is the standard Euclidean metric defined on $\mathbb{R}$, i.e., $d_1(x, y) := |y - x|$. Also, let $\mathbb{D}_n = \mathbb{D} = \mathbb{D}([0, M])$ for each $n \in \mathbb{N}$ and $\mathbb{D}_0 = \mathbb{C}([0, M])$. Since $\mathbb{C}([0, M])$ is separable as a subspace of the Skorokhod space $\mathbb{D}([0, M])$, it turns out that $W(\mathbf{v}; n) : \Omega \to \mathbb{D}([0, M])$ and $W : \Omega \to \mathbb{D}([0, M])$ satisfy the conditions (i) and (ii) for Lemma A.8. Thus, it only remains to verify the condition (iii) of Lemma A.8. To this end, we need to show that for any sequence $(x_n : n \in \mathbb{N})$ in $\mathbb{D}([0, M])$ such that $x_n \to x$ in $\mathbb{D}([0, M])$ as $n \to \infty$ together with $x \in \mathbb{C}([0, M])$, we have

$$\lim_{n \to \infty} F_n(x_n) = F(x). \tag{E.59}$$

Let us write

$$\begin{aligned} F_n(x_n) - F(x) = &\left[\frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x_n\left(\frac{t}{n}\right)\right\} - \frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x\left(\frac{t}{n}\right)\right\}\right] \\ &+ \left[\frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x\left(\frac{t}{n}\right)\right\} - \int_0^M \exp\{x(t)\}\, \mathrm{d}t\right], \end{aligned}$$

so that one can obtain

$$\begin{aligned} |F_n(x_n) - F(x)| \leq &\underbrace{\left|\frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x_n\left(\frac{t}{n}\right)\right\} - \frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x\left(\frac{t}{n}\right)\right\}\right|}_{=: \mathcal{I}_n} \\ &+ \underbrace{\left|\frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor} \exp\left\{x\left(\frac{t}{n}\right)\right\} - \int_0^M \exp\{x(t)\}\, \mathrm{d}t\right|}_{=: \mathcal{J}_n}. \end{aligned} \tag{E.60}$$

Based on the assumption that $x_n \to x$ in $\mathbb{D}([0,M])$ as $n \to \infty$ and $x \in \mathbb{C}([0,M])$, we also obtain the uniform convergence:

$$\|x_n - x\|_{[0,M]} = \sup\{|x_n(t) - x(t)| : t \in [0,M]\} \to 0 \tag{E.61}$$

as $n \to \infty$.

We now control the first term $\mathcal{I}_n$. From the continuity of $x : [0,M] \to \mathbb{R}$, there exists some universal constant $C \in (0,+\infty)$ such that $\|x\|_{[0,M]} \leq C$. Due to the uniform convergence equation E.61, there is some $N \in \mathbb{N}$ such that

$$\|x_n\|_{[0,M]} \leq \|x_n - x\|_{[0,M]} + \|x\|_{[0,M]} \leq 1 + C, \quad n \geq N.$$

Thus, all values of $x_n(t)$ and $x(t)$ lie in the compact interval $[-1-C, 1+C]$ for all $t \in [0,M]$ if $n \geq N$. On the other hand, the uniform continuity of the exponential function $\exp : [-1-C, 1+C] \to (0,+\infty)$ gives

$$\lim_{n\to\infty} \sup\{|\exp\{x_n(t)\} - \exp\{x(t)\}| : t \in [0,M]\} = 0. \tag{E.62}$$

Therefore, it follows from the fact equation E.62 that

$$\mathcal{I}_n \leq \frac{\lfloor nM \rfloor + 1}{n} \cdot \sup\{|\exp\{x_n(t)\} - \exp\{x(t)\}| : t \in [0,M]\} \to 0 \tag{E.63}$$

as $n \to \infty$.

Next, let us control the second term $\mathcal{J}_n$. The continuity of $x : [0,M] \to \mathbb{R}$ implies the continuity of $\exp \circ x : [0,M] \to (0,+\infty)$, thereby it is Riemann integrable. Therefore, its left-endpoint Riemann sum converges:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor - 1} \exp\left\{x\left(\frac{t}{n}\right)\right\} = \int_0^M \exp\{x(t)\}\,\mathrm{d}t. \tag{E.64}$$

Hence, we obtain

$$\mathcal{J}_n \leq \underbrace{\left|\frac{1}{n} \sum_{t=0}^{\lfloor nM \rfloor - 1} \exp\left\{x\left(\frac{t}{n}\right)\right\} - \int_0^M \exp\{x(t)\}\,\mathrm{d}t\right|}_{\to 0} + \underbrace{\frac{1}{n} \exp\left\{x\left(\frac{\lfloor nM \rfloor}{n}\right)\right\}}_{\to 0 \cdot \exp\{x(M)\} = 0} \to 0 \tag{E.65}$$

as $n \to \infty$ by using the observation equation E.64. Taking two conclusions equation E.63 and equation E.65 collectively completes the verification of the condition (iii) of Lemma A.8 by appealing to the inequality equation E.60.

Finally, it is time to put everything together. By applying the extended continuous mapping theorem (Lemma A.8) together with the functional CLT for $W(\mathbf{v}; n)$ (Lemma E.3), one can reveal that

$$\mathcal{C}_n(\mathbf{v}; M) = F_n\{W(\mathbf{v}; n)\} \Rightarrow F(W) = \mathcal{C}(M)$$

as $n \to \infty$, concluding the proof of Lemma E.4.

□

Our next objective is to prove that the contribution of the indices $t > \lfloor nM \rfloor$ to the sum $\frac{\mathfrak{R}(\mathbf{v};n)}{n} = \frac{1}{n}\sum_{t=0}^{\infty} \exp\{S_t(\mathbf{v}; n)\}$ vanishes as $M \to +\infty$ uniformly in n. To this end, we define the tail part

$$\mathcal{T}_n(\mathbf{v}; M) := \frac{\mathfrak{R}(\mathbf{v}; n)}{n} - \mathcal{C}_n(\mathbf{v}; M) = \frac{1}{n} \sum_{t=\lfloor nM \rfloor + 1}^{\infty} \exp\{S_t(\mathbf{v}; n)\} \tag{E.66}$$

for each fixed $M \in (0,+\infty)$.

Lemma E.5 (Uniform tail bound in expectation)**.** *It holds for any* $(M, n) \in (0, +\infty) \times (\mathbb{N} \setminus \{1\})$ *that*

$$\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\mathcal{T}_n \left(\mathbf{v}; M \right) \right] \leq \exp\left(-M\right). \tag{E.67}$$

Consequently, it is immediate from Markov's inequality that

$$\sup \left\{ \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left\{ \mathcal{T}_n \left(\mathbf{v}; M \right) > \varepsilon \right\} : n \in \mathbb{N} \setminus \{1\} \right\} \leq \frac{\exp\left(-M\right)}{\varepsilon} \to 0 \tag{E.68}$$

as $M \to +\infty$*, for every fixed* $\varepsilon \in (0, +\infty)$.

Proof of Lemma E.5. Since $\exp\left\{ S_0 \left(\mathbf{v}; n \right) \right\} = 1$ and $\exp\left\{ S_t \left(\mathbf{v}; n \right) \right\} = \prod_{j=0}^{t-1} \left\{ \frac{Z_j(\mathbf{v}; n)}{n} \right\}$ for $t \in \mathbb{N}$ together with the fact equation E.10 that $\left\{ Z_t \left(\mathbf{v}; n \right) : t \in \mathbb{Z}_+ \right\}$ are i.i.d. random variables with the χ^2-square distribution with $n - 1$ degrees of freedom, we obtain

$$\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\exp\left\{ S_t \left(\mathbf{v}; n \right) \right\} \right] = \left\{ \frac{\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[Z_0(n) \right]}{n} \right\}^t = \left(1 - \frac{1}{n} \right)^t, \quad t \in \mathbb{Z}_+. \tag{E.69}$$

Hence, the monotone convergence theorem together with the finding equation E.69 yields

$$\begin{aligned}
\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\mathcal{T}_n \left(\mathbf{v}; M \right) \right] &= \frac{1}{n} \sum_{t=\lfloor nM \rfloor + 1}^{\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\exp\left\{ S_t \left(\mathbf{v}; n \right) \right\} \right] \\
&= \frac{1}{n} \sum_{t=\lfloor nM \rfloor + 1}^{\infty} \left(1 - \frac{1}{n} \right)^t \\
&= \left(1 - \frac{1}{n} \right)^{\lfloor nM \rfloor + 1} \\
&\leq \exp\left(-M\right)
\end{aligned}$$

as desired in the inequality equation E.67.

□

Lastly, it is time to put everything together to complete the proof of Theorem E.1. Given any horizon $M \in (0, +\infty)$, we may write

$$\frac{\mathfrak{R}\left(\mathbf{v}; n \right)}{n} = \mathcal{C}_n \left(\mathbf{v}; M \right) + \mathcal{T}_n \left(\mathbf{v}; M \right), \quad n \geq 2. \tag{E.70}$$

We first claim that $\mathfrak{I} := \int_0^\infty \exp\left\{ \sqrt{2}\mathbb{B}(t) - 2t \right\} \mathrm{d}t = \int_0^\infty \exp\left\{ W(t) \right\} \mathrm{d}t < +\infty$ $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost surely. Since we know that $\lim_{t \to \infty} \frac{\mathbb{B}(t)}{t} = 0$ $\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost surely, we obtain

$$\frac{W(t)}{t} = \sqrt{2} \cdot \frac{\mathbb{B}(t)}{t} - 2 \to -2 \tag{E.71}$$

$\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}$-almost surely as $t \to \infty$. Let $\mathfrak{E}$ denote the event for which equation E.71 holds. Then for each $\omega \in \mathfrak{E}$, there exists some $t_0(\omega) \in \mathbb{R}_+$ such that $W\left(\omega, t\right) \leq -t$ for all $t \in [t_0(\omega), +\infty)$. Thus, one can reveal that

$$\begin{aligned}
\int_0^\infty \exp\left\{ W\left(\omega, t\right) \right\} \mathrm{d}t &= \int_0^{t_0(\omega)} \exp\left\{ W\left(\omega, t\right) \right\} \mathrm{d}t + \int_{t_0(\omega)}^\infty \exp\left\{ W\left(\omega, t\right) \right\} \mathrm{d}t \\
&\leq \int_0^{t_0(\omega)} \exp\left\{ W\left(\omega, t\right) \right\} \mathrm{d}t + \underbrace{\int_{t_0(\omega)}^\infty \exp\left(-t\right) \mathrm{d}t}_{= \exp\{-t_0(\omega)\}} \\
&< +\infty,
\end{aligned}$$

which implies $\mathfrak{E} \subseteq \left\{ \int_0^\infty \exp\left\{ W(t) \right\} \mathrm{d}t < +\infty \right\}$. Hence, one may conclude that

$$\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left\{ \mathfrak{I} = \int_0^\infty \exp\left\{ W(t) \right\} \mathrm{d}t < +\infty \right\} = 1. \tag{E.72}$$

In particular, since $\exp\{W(t)\} > 0$ for every $t \in \mathbb{R}_+$, we have $\mathcal{C}(M) \uparrow \mathfrak{I}$ as $M \to +\infty$.

On the other hand, the uniform negligibility of the truncation error $\mathcal{T}_n(\mathbf{v}; M)$ equation E.68 together with the decomposition equation E.70 gives for every $\varepsilon > 0$ that

$$\begin{aligned} &\sup\left\{\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left\{\left|\frac{\mathfrak{R}(\mathbf{v}; n)}{n} - \mathcal{C}_n(\mathbf{v}; M)\right| > \varepsilon\right\} : n \in \mathbb{N} \setminus \{1\}\right\} \\ &= \sup\left\{\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left\{\mathcal{T}_n(\mathbf{v}; M) > \varepsilon\right\} : n \in \mathbb{N} \setminus \{1\}\right\} \to 0 \end{aligned} \tag{E.73}$$

as $M \to +\infty$. In order to establish the final goal $\frac{\mathfrak{R}(\mathbf{v};n)}{n} \xrightarrow{d} \mathfrak{I}$ as $n \to \infty$, it is sufficient to prove that

$$\lim_{n\to\infty} \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}(\mathbf{v}; n)}{n}\right\}\right] = \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f(\mathfrak{I})\right] \tag{E.74}$$

for any bounded and uniformly continuous function $f : \mathbb{R} \to \mathbb{R}$ by virtue of *Theorem 11.3.3* in Dudley (2002). Towards this end, let us fix any bounded and uniformly continuous function $f : \mathbb{R} \to \mathbb{R}$ and consider its *modulus of continuity* $\omega_f : (0, +\infty) \to \mathbb{R}_+$:

$$\omega_f(\delta) := \sup\left\{|f(x) - f(y)| : (x, y) \in \mathbb{R} \times \mathbb{R} \text{ such that } |x - y| \leq \delta\right\}, \quad \delta \in (0, +\infty).$$

Due to the uniform continuity of $f : \mathbb{R} \to \mathbb{R}$, we have

$$\lim_{\delta \downarrow 0^+} \omega_f(\delta) = 0. \tag{E.75}$$

Now, we observe for any $\delta \in (0, +\infty)$ that

$$|f(x) - f(y)| \leq \omega_f(\delta) \cdot \mathbb{1}_{\{|x-y| \leq \delta\}} + 2\|f\|_\infty \cdot \mathbb{1}_{\{|x-y| > \delta\}}, \quad (x, y) \in \mathbb{R} \times \mathbb{R}.$$

which immediately deduces

$$\mathbb{E}\left[|f(X) - f(Y)|\right] \leq \omega_f(\delta) + 2\|f\|_\infty \cdot \mathbb{P}\left\{|x - y| > \delta\right\} \tag{E.76}$$

for any real-valued random variables $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. By applying the inequality equation E.76 with $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\mathbb{R}^d \times \{\pm 1\}, \mathcal{B}\left(\mathbb{R}^d\right) \otimes \mathscr{P}\left(\{\pm 1\}\right), \mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\right)$ and $(X, Y) = \left(\frac{\mathfrak{R}(\mathbf{v};n)}{n}, \mathcal{C}_n(\mathbf{v}; M)\right)$, we now reach

$$\begin{aligned} &\sup\left\{\left|\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}(\mathbf{v}; n)}{n}\right\}\right] - \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\mathcal{C}_n(\mathbf{v}; M)\right\}\right]\right| : n \in \mathbb{N} \setminus \{1\}\right\} \\ &\leq \omega_f(\delta) + 2\|f\|_\infty \sup\left\{\mathbb{P}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left\{\left|\frac{\mathfrak{R}(\mathbf{v}; n)}{n} - \mathcal{C}_n(\mathbf{v}; M)\right| > \varepsilon\right\} : n \in \mathbb{N} \setminus \{1\}\right\}. \end{aligned} \tag{E.77}$$

By letting $M \to +\infty$ in the inequality equation E.77, we obtain

$$\begin{aligned} &\limsup_{M\to+\infty} \sup\left\{\left|\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}(\mathbf{v}; n)}{n}\right\}\right] - \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\mathcal{C}_n(\mathbf{v}; M)\right\}\right]\right| : n \in \mathbb{N} \setminus \{1\}\right\} \\ &\leq \omega_f(\delta) \xrightarrow{\delta \downarrow 0^+} 0, \end{aligned}$$

which tells us that

$$\limsup_{M\to+\infty} \sup_{n \in \mathbb{N}\setminus\{1\}} \underbrace{\left|\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}(\mathbf{v}; n)}{n}\right\}\right] - \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\mathcal{C}_n(\mathbf{v}; M)\right\}\right]\right|}_{=: \mathscr{A}_n(\mathbf{v}; M, f)} = 0. \tag{E.78}$$

On the other hand, we apply the bounded convergence theorem to obtain

$$\kappa(\mathbf{v}; M, f) := \left|\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\mathcal{C}(M)\right\}\right] - \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left(\mathfrak{I}\right)\right]\right| \to 0 \tag{E.79}$$

as $M \to +\infty$, since $\mathcal{C}(M) \uparrow \mathfrak{I}$ as $M \to +\infty$. Also, Lemma E.4 together with the continuous mapping theorem yields for any horizon $M \in (0, +\infty)$ that

$$\mathscr{B}_n(\mathbf{v}; M, f) := \left|\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\mathcal{C}_n(\mathbf{v}; M)\right\}\right] - \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)}\left[f\left\{\mathcal{C}(M)\right\}\right]\right| \to 0 \tag{E.80}$$

as $n \to \infty$. By appealing to the triangle inequality, we get

$$\begin{aligned}&\left|\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}\left(\mathbf{v};n\right)}{n}\right\}\right]-\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f(\mathfrak{I})\right]\right| \\ &\leq \mathscr{A}_n\left(\mathbf{v};M,f\right)+\mathscr{B}_n\left(\mathbf{v};M,f\right)+\kappa\left(\mathbf{v};M,f\right).\end{aligned} \tag{E.81}$$

Take $\limsup_{n\to\infty}$ to the bound equation E.81 to reach

$$\begin{aligned}&\limsup_{n\to\infty}\left|\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}\left(\mathbf{v};n\right)}{n}\right\}\right]-\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f(\mathfrak{I})\right]\right| \\ &\leq \sup_{n\in\mathbb{N}\setminus\{1\}}\mathscr{A}_n\left(\mathbf{v};M,f\right)+\kappa\left(\mathbf{v};M,f\right).\end{aligned} \tag{E.82}$$

due to the observation equation E.80, for every fixed horizon $M \in (0, +\infty)$. By letting $M \to +\infty$ in the inequality equation E.82, one has

$$\lim_{n\to\infty}\left|\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f\left\{\frac{\mathfrak{R}\left(\mathbf{v};n\right)}{n}\right\}\right]-\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f(\mathfrak{I})\right]\right|=0$$

by virtue of the facts equation E.78 and equation E.79, which proves the desired claim equation E.74. This concludes the proof of Theorem E.1.

E.3 Proof of Theorem E.2

We embark on the proof of Theorem E.2 with the introduction of the continuous-time Markov process $W^x : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ defined by

$$W^x\left(\omega,t\right) := x + W\left(\omega,t\right) = x + \sqrt{2}\mathbb{B}\left(\omega,t\right) - 2t, \tag{E.83}$$

started from $W^x(0) = x$ for any given $x \in \mathbb{R}$. From the literature of stochastic processes and Markov semi-group theory, the *infinitesimal generator* $\mathcal{L} : \mathcal{C}^2(\mathbb{R}) \to (\mathbb{R} \to \mathbb{R})$ of $W^x : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is

$$\mathcal{L}(f)(x) := \lim_{t\downarrow 0^+}\frac{\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f\left\{W^x(t)\right\}\right]-f(x)}{t}, \quad f \in \mathcal{C}^2(\mathbb{R}), \tag{E.84}$$

where $\mathcal{C}^2(\mathbb{R})$ denotes the space of all real-valued twice continuously differentiable functions defined on the real line $\mathbb{R}$. Here, we provide the explicit formula for the infinitesimal generator $\mathcal{L} : \mathcal{C}^2(\mathbb{R}) \to (\mathbb{R} \to \mathbb{R})$ of $W^x : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ through the following lemma.

Lemma E.6. *For any $f \in \mathcal{C}^2(\mathbb{R})$, it holds that*

$$\mathcal{L}(f)(x) = f''(x) - 2f'(x), \quad x \in \mathbb{R}. \tag{E.85}$$

Proof of Lemma E.6. Leveraging the second-order Taylor expansion of $f \in \mathcal{C}^2(\mathbb{R})$ centered at $x \in \mathbb{R}$, we obtain

$$f\left\{W^x(t)\right\} = f\left\{x + W(t)\right\} = f(x) + f'(x)W(t) + \frac{1}{2}f''(x)W^2(t) + o\left\{W^2(t)\right\}. \tag{E.86}$$

By taking expectation to the equation equation E.86, we obtain from $W(t) = \sqrt{2}\mathbb{B}(t) - 2t \sim \mathcal{N}\left(-2t, 2t\right)$ that

$$\begin{aligned}\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[f\left\{W^x(t)\right\}\right] &= f(x) + f'(x)\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[W(t)\right] + \frac{1}{2}f''(x)\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[W^2(t)\right] + o(t) \\ &= f(x) - 2tf'(x) + \frac{1}{2}f''(x)\left(2t + 4t^2\right) + o(t) \\ &= f(x) - 2tf'(x) + tf''(x) + o(t)\end{aligned} \tag{E.87}$$

as $t \downarrow 0^+$. Hence, it follows from equation E.87 that

$$\frac{\mathbb{E}\left[f\left\{W^x(t)\right\}\right]-f(x)}{t} = -2f'(x) + f''(x) + o(1) \to -2f'(x) + f''(x)$$

as $t \downarrow 0^+$, concluding the proof of Lemma E.6.

□

Let us define $u_q : \mathbb{R} \to \mathbb{R}$ for $q \in \mathbb{R}_+$ as

$$\begin{aligned} u_q(x) &:= \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\exp\left\{ -q \int_0^\infty \exp\left(W^x(t)\right) \mathrm{d}t \right\} \right] \\ &= \exp\left(-qx\right) \mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\exp\left(-q\mathfrak{I}\right) \right]. \end{aligned} \tag{E.88}$$

By the definition, it is clear that the Laplace transform of the weak limit $\mathfrak{I}$ is given by

$$\mathbb{E}_{(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star)} \left[\exp\left(-q\mathfrak{I}\right) \right] = u_q(0). \tag{E.89}$$

Due to the Feynman-Kac formula Karatzas & Shreve (1991), one can find that $u_q \in \mathcal{C}^2(\mathbb{R})$ and it solves following the partial differential equation (PDE):

$$0 = \mathcal{L}\left(u_q\right)(x) - q \exp(x) u_q(x) \overset{\text{(a)}}{=} u_q''(x) - 2u_q'(x) - q \exp(x) u_q(x), \tag{E.90}$$

whether the step (a) follows from Lemma E.6, together with the boundary conditions:

$$u_q\left(-\infty\right) := \lim_{x \to -\infty} u_q(x) = 1 \quad \text{and} \quad u_q\left(+\infty\right) := \lim_{x \to +\infty} u_q(x) = 0. \tag{E.91}$$

We next turn to solve the PDE equation E.90 with the boundary conditions equation E.91. In order to do so, we assume that $q > 0$ and define a smooth bijection $\theta : (0, +\infty) \to \mathbb{R}$ by

$$x = \theta_q(z) := 2 \log\left(\frac{z}{2\sqrt{q}}\right) \quad \Leftrightarrow z = \theta_q^{-1}(x) = 2\sqrt{q} \exp\left(\frac{x}{2}\right).$$

By letting $v_q : (0, +\infty) \to \mathbb{R}$, where

$$v_q(z) := \exp\left\{-\theta_q(z)\right\} u_q\left\{\theta_q(z)\right\}, \quad z \in (0, +\infty),$$

one can reveal that $v_q \in \mathcal{C}^2(\mathbb{R})$ and it solve the following *modified Bessel equation of order 2*:

$$z^2 v_q''(z) + z v_q'(z) - \left(z^2 + 4\right) v_q(z) = 0. \tag{E.92}$$

Therefore, the solution to the modified Bessel equation equation E.92 of order 2 is given by

$$v_q(z) = \alpha(q) I_2(z) + \beta(q) K_2(z), \quad z \in (0, +\infty), \tag{E.93}$$

for some constants $\alpha(q) \in \mathbb{R}$ and $\beta(q) \in \mathbb{R}$, where $I_2 : (0, +\infty) \to \mathbb{R}$ and $K_2 : (0, +\infty) \to \mathbb{R}$ define the *order-2 modified Bessel functions* of the first and second kind, respectively:

$$I_2(z) := \frac{z^2}{8} \sum_{k=0}^{\infty} \frac{1}{k!(k+2)!} \left(\frac{z^2}{4}\right)^k, \quad K_2(z) := \frac{z^2}{8} \int_0^\infty t^{-3} \exp\left(-t - \frac{z^2}{4t}\right) \mathrm{d}t. \tag{E.94}$$

Hence, taking the general solution equation E.93 into the identity $u_q(x) = \exp(x) v_q\left\{\theta_q^{-1}(x)\right\}$ gives

$$u_q(x) = \exp(x) \left[\alpha(q) I_2\left\{ 2\sqrt{q} \exp\left(\frac{x}{2}\right) \right\} + \beta(q) K_2\left\{ 2\sqrt{q} \exp\left(\frac{x}{2}\right) \right\} \right], \quad x \in \mathbb{R}. \tag{E.95}$$

Now, we compare the general form equation E.94 of $u_q : \mathbb{R} \to \mathbb{R}$ together with the boundary condition equation E.91 in order to specify the numerical constants $\alpha(q) \in \mathbb{R}$ and $\beta(q) \in \mathbb{R}$. To this end, we recall the following asymptotic behaviors of the order-2 modified Bessel functions:

- As $z \downarrow 0^+$, we have

$$I_2(z) \sim \frac{z^2}{8} \quad \text{and} \quad K_2(z) \sim \frac{2}{z^2}. \tag{E.96}$$

- As $z \to +\infty$, we have

$$I_2(z) \sim \frac{\exp(z)}{\sqrt{2\pi z}} \quad \text{and} \quad K_2(z) \sim \sqrt{\frac{\pi}{2z}} \exp\left(-z\right). \tag{E.97}$$

We first take $x \to +\infty$ to equation E.94. Since the boundary condition equation E.91 requires $\lim_{x\to+\infty} u_q(x) = 0$ and we have from the asymptotic behaviors equation E.97 that

$$\lim_{x\to+\infty} \exp(x) \cdot I_2\left\{2\sqrt{q}\exp\left(\frac{x}{2}\right)\right\} = +\infty \quad \text{and} \quad \lim_{x\to+\infty} \exp(x) \cdot K_2\left\{2\sqrt{q}\exp\left(\frac{x}{2}\right)\right\} = 0,$$

we must have $\alpha(q) = 0$. On the other hand, we take $x \downarrow 0^+$ to the equation equation E.94. As the boundary condition equation E.91 requires $\lim_{x\downarrow 0^+} u_q(x) = 1$ and one can find by utilizing the asymptotic behaviors equation E.96 that

$$\lim_{x\downarrow 0^+} \exp(x) \cdot K_2\left\{2\sqrt{q}\exp\left(\frac{x}{2}\right)\right\} = \frac{1}{2q},$$

we must choose $\beta(q) = 2q$. As a consequence, the unique solution to the PDE equation E.90 with boundary conditions equation E.91 is

$$u_q(x) = 2q\exp(x) \cdot K_2\left\{2\sqrt{q}\exp\left(\frac{x}{2}\right)\right\}, \quad x \in \mathbb{R}. \tag{E.98}$$

In particular, by putting $x = 0$ into equation E.98, the Laplace transform of the weak limit $\mathfrak{I}$ becomes

$$\mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left(-q\mathfrak{I}\right)\right] = u_q(0) = 2qK_2\left(2\sqrt{q}\right), \quad q \in (0,+\infty). \tag{E.99}$$

Lastly, it is time to wrap up the proof for Theorem E.2. If $G \sim \Gamma(2,1)$, then its density function is given by $f_G(g) = g\exp\left(-g\right)\mathbb{1}_{(0,+\infty)}(g)$. Thus, the Laplace transform of $1/G$ is

$$\mathbb{E}_{G\sim\Gamma(2,1)}\left[\exp\left(-\frac{q}{G}\right)\right] = \int_0^\infty g\exp\left(-g-\frac{q}{g}\right)\mathrm{d}g, \quad q \in \mathbb{R}_+. \tag{E.100}$$

Here, we make use of the standard Bessel identity: for any $(\beta,\gamma,\nu) \in (0,+\infty)\times(0,+\infty)\times(0,+\infty)$,

$$\int_0^\infty x^{\nu-1}\exp\left(-\beta x-\frac{\gamma}{x}\right)\mathrm{d}x = 2\left(\frac{\gamma}{\beta}\right)^{\frac{\nu}{2}} K_\nu\left(2\sqrt{\beta\gamma}\right). \tag{E.101}$$

By taking $(\beta,\gamma,\nu) = (1,q,2)$ in the identity equation E.101 and then combining the result together with the equations equation E.99 and equation E.100, one has

$$\mathbb{E}_{G\sim\Gamma(2,1)}\left[\exp\left(-\frac{q}{G}\right)\right] = 2qK_2\left(2\sqrt{q}\right) = \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left(-q\mathfrak{I}\right)\right], \quad q \in (0,+\infty). \tag{E.102}$$

Hence, it follows that

$$\mathbb{E}_{G\sim\Gamma(2,1)}\left[\exp\left(-\frac{q}{G}\right)\right] = \mathbb{E}_{(\boldsymbol{\mu}_\star,\boldsymbol{\Sigma}_\star)}\left[\exp\left(-q\mathfrak{I}\right)\right], \quad q \in \mathbb{R}_+, \tag{E.103}$$

which implies $\mathfrak{I} \overset{d}{=} \frac{1}{G}$ for $G \sim \Gamma(2,1)$ due to the classical uniqueness theorem for Laplace transforms. This completes the proof of Theorem E.2.

F Omitted details for Section 6

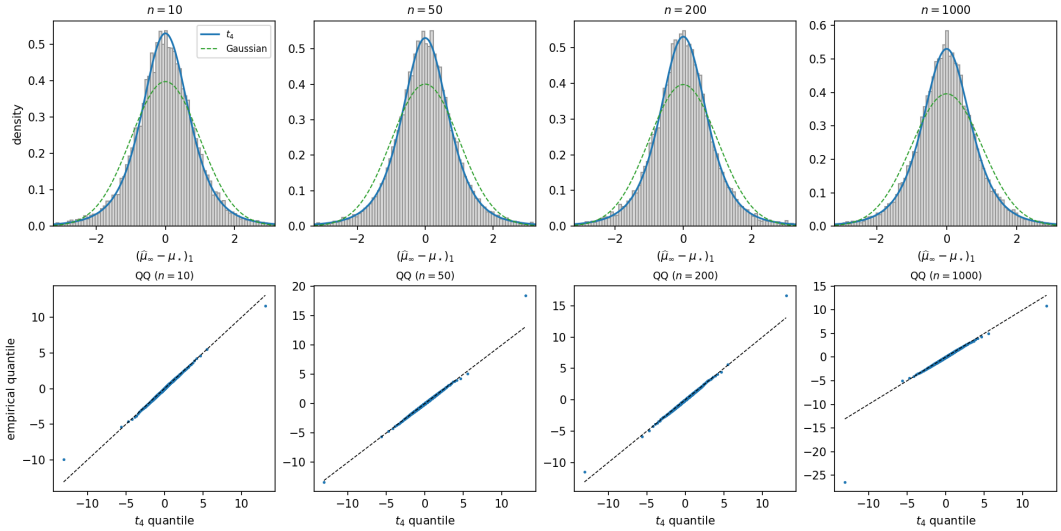


Figure 2: Distributional limit of the terminal mean vector $\widehat{\boldsymbol{\mu}}_\infty(n)$. **Top**: the marginal histograms of $\mathbf{e}_1^\top \{\widehat{\boldsymbol{\mu}}_T(n) - \boldsymbol{\mu}_\star\}$ with the density function of $t_1\,(0, 1/2, 4)$ (solid) and the normal density function of matching SD (dashed) overlaid. **Bottom**: QQ plots versus the $t_1\,(0, 1/2, 4)$-quantiles.