ML ESTIMATION FROM BITS

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating statistical parameters from quantized signals has received significant attention in recent years, as recovering information from quantized measurements has numerous applications across signal processing, communications, and data analysis. In this work, we focus on maximum likelihood (ML) estimation of statistical parameters from quantized samples. Directly solving the ML problem is challenging, as the likelihood function involves multiple integrals that are difficult to evaluate. To address this challenge, we propose an expectation-conditional-maximization (ECM) algorithm under a general distributional framework. Our approach generalizes the quantization model to multi-bit settings and allows the underlying signal to follow any distribution within the normal mean-variance mixture family. By designing suitable surrogate functions, the ECM algorithm ensures that all model parameters can be updated in closed form at each iteration. Leveraging the ECM framework, we provide convergence guarantees, and under specific distributional assumptions, we further derive bounds on the convergence rate and the statistical error. Extensive experiments demonstrate the effectiveness of our method in recovering statistical parameters from quantized data.

1 INTRODUCTION

Quantization, which represents signals using a finite number of bits, offers a hardware-efficient approach for data storage and transmission, reducing power consumption while maintaining acceptable accuracy (Roberts, 1962; Widrow & Kollár, 2008). These properties have led to the widespread adoption of quantization in applications that require efficient processing of large-scale data.

In recommendation systems, for example, user ratings are often quantized (e.g., binary preferences or 1-5 star ratings), motivating extensive research on recovering the underlying rating matrix (Davenport et al., 2014; Cai & Zhou, 2013; Bhaskar, 2016; Bottegal & Suykens, 2017; Gao et al., 2018). In compressed sensing, quantization arises when estimating sparse signals from limited measurements, with a growing focus on one-bit and multi-bit quantized sensing (Boufounos & Baraniuk, 2008a; Zymnis et al., 2009; Plan & Vershynin, 2013; Ai et al., 2014; Shao et al., 2024). Quantization is also widely used in wireless communication and sensing, including channel estimation, array detection, and radar signal processing (Choi et al., 2016; Stöckle et al., 2016; Plabst et al., 2018; Ren & Li, 2017; Ameri et al., 2019; Jin et al., 2020; Bar-Shalom & Weiss, 2002b; Lu et al., 2024). Beyond signal recovery, quantized data naturally appear in regression and subspace learning, giving rise to quantized regression and subspace estimation problems (Mayne, 1967; Gyorfi & Wegkamp, 2008; Chen et al., 2023; Chi & Fu, 2017; Dirksen et al., 2025).

In many applications, the primary interest is not the signal itself, but its underlying statistical properties. Researchers accomplish downstream tasks such as classification and anomaly detection by estimating the mean and covariance matrix of the signal. In many applications, the primary interest is not the signal itself, but its underlying statistical properties, such as the mean (Papadopoulos et al., 2001; Dabeer & Masry, 2008) for distributed detection (Ribeiro & Giannakis, 2006a;b; Fang & Li, 2008), and the covariance matrix (Van Vleck & Middleton, 1966; Gray & Stockham, 1993; Liu & Lin, 2021; Dirksen et al., 2022; Xiao et al., 2023) for direction-of-arrival estimation (Eamaz et al., 2022; 2023; Bar-Shalom & Weiss, 2002a), power estimation in wireless sensor networks (Mo et al., 2017), spectrum sensing (Yang et al., 2025), and networked sensing (Chi & Fu, 2017).

However, quantization inevitably discards part of the original information, making the recovery of signals or their statistical properties from quantized measurements a challenging and practically

important problem. In this paper, we study parameter estimation of $\boldsymbol{x}$ from quantized measurements. Given a random signal $\boldsymbol{x}$, the quantized measurement $\boldsymbol{y}$ is obtained in the following way

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{x}), \tag{1}$$

where $\mathcal{Q} : \mathbb{R}^d \to \mathcal{K}^d$ is a quantization function, or quantizer, with the set $\mathcal{K}$ containing finite e elements, i.e., $\mathcal{K} = \{k_1, \ldots, k_e\}$. The definition of the i-th element $[\mathcal{Q}(\boldsymbol{a})]_i$ is given by

$$[\mathcal{Q}(\boldsymbol{a})]_i = k_l, \quad \text{if } a_i \in [\tau_{l-1}, \tau_l), \tag{2}$$

where $\bigcup_{l=1}^{e}[\tau_{l-1}, \tau_l) = \mathbb{R}$ with $\tau_0 = -\infty$ and $\tau_e = \infty$. When the number of bits used to represent the quantized value, i.e., $\log_2 e$, is small, we call it a coarse quantization. The coarsest quantization method, which is also the most commonly adopted one in practice, is the one-bit quantization, i.e., $e = 2$. With the additional conditions $k_1 = -1$, $k_2 = 1$, and $\tau_1 = 0$, the quantization function $\mathcal{Q}$ becomes the element-wise signum function.

1.1 Related works

Parameter estimation from quantized signals has been extensively studied. Among the many directions, two fundamental lines of inquiry concern the estimation of the mean and the covariance of $\boldsymbol{x}$, which we take as our starting point. Furthermore, we will demonstrate that our proposed framework is versatile and can be extended to various related tasks, including quantized regression, quantized matrix completion, and quantized compressed sensing.

Quantized mean estimation In Papadopoulos et al. (2001), in order to estimate signals from wireless sensor networks, an optimization problem is formulated for estimating the mean under a one-dimensional Gaussian distribution assumption, based on multi-bit quantized data. To address this optimization problem, the authors propose an ECM-based algorithm. Subsequently, Ribeiro & Giannakis (2006b) focuses on the specific scenario of one-bit quantization under a Gaussian distribution and derives a closed-form expression for mean estimation in this case. Further, Fang & Li (2008) extends this line of research by introducing an adaptive threshold into the one-bit quantization function, thereby improving mean estimation accuracy. In addition, Dabeer & Masry (2008) generalizes the problem to the multidimensional setting. Ribeiro & Giannakis (2006a) extends the model distribution to the generalized Gaussian distribution under the one-dimensional assumption.

Quantized covariance/correlation estimation In Van Vleck & Middleton (1966), the authors investigate the estimation of a correlation matrix from one-bit quantized zero-mean Gaussian measurements, based on the arcsine law (Lévy, 1940). However, in such one-bit zero-mean settings, the individual variance of each dimension cannot be determined, which prevents recovery of the full covariance matrix. To estimate the variance, the "dithering technique" is introduced, which is to set a threshold in the signum function (Liu & Lin, 2021). Based on the dithering technique, the variance can be obtained in closed form (Fang & Li, 2008), but the correlation matrix should be solved analytically. Consequently, subsequent studies (Dirksen et al., 2022; Eamaz et al., 2023; Xiao et al., 2023; Liu & Chou, 2025) have proposed various strategies for correlation estimation. The estimation approaches for correlation can be classified into two categories. The first category relies on the correlation coefficient function of the one-bit Gaussian distribution, with analyses usually restricted to pairwise interactions between dimensions in the multivariate setting. Since this function is generally computationally intractable, different approximate functions have been proposed to compute the correlation coefficient (Eamaz et al., 2023; Xiao et al., 2023; Liu & Chou, 2025). In Eamaz et al. (2022; 2023), the authors introduce a known, time-varying threshold for the signum function and propose a modified arcsine law to express correlation coefficients as integrals, which are then approximated using Gauss–Legendre quadratures. The method in (Xiao et al., 2023) applies maximum likelihood (ML) estimation to compute correlation coefficients, which is equivalent to iteratively evaluating their approximate functions. The work (Liu & Chou, 2025) represents the correlation coefficient function as an infinite series via the one-bit Hermite law (Liu & Lin, 2021) and then approximates it using harmonic approximation. The second category (Dirksen et al., 2022; Chen et al., 2024) directly employs the sample covariance matrix computed using multiple thresholds. These methods are not restricted to one-bit quantization and can be generalized to multi-bit cases, though typically at the expense of reduced estimation accuracy.

1.2 CONTRIBUTION

In this paper, we study the maximum likelihood (ML) estimation of the parameters of a random signal $\boldsymbol{x}$ with quantization. We assume that $\boldsymbol{x}$ follows a general normal mean–variance mixture model (Barndorff-Nielsen et al., 1982), which is a flexible family encompassing several well-known distributions, including Gaussian, t, generalized hyperbolic skew-t (GHST), hyperbolic, and generalized hyperbolic (GH) distributions. Unlike previous approaches that are restricted to one-bit or Gaussian settings, our framework accommodates arbitrary bit quantization functions and any distribution within the normal mean–variance mixture family. To solve the ML estimation problem, we develop an expectation conditional maximization (ECM) algorithm. The method alternates between two steps: (i) an expectation step (E-step), where we construct a surrogate of the likelihood via Jensen's inequality, and (ii) a conditional maximization step (CM-step), where the surrogate is maximized with respect to a subset of parameters while holding the others fixed. This procedure yields closed-form updates for the location, skewness, and scatter parameters. The proposed ECM algorithm inherits the interpretability and stability of the ECM framework, while we further establish that it converges globally at a linear rate.

2 PROBLEM FORMULATION

A random variable $\boldsymbol{x} \in \mathbb{R}^d$ following a normal mean-variance mixture model is represented by

$$\boldsymbol{x} = \boldsymbol{\mu} + z\boldsymbol{\xi} + (z\boldsymbol{\Sigma})^{\frac{1}{2}} \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{\mu}$ is the location parameter, $\boldsymbol{\xi}$ is the skewness parameter, $\boldsymbol{\Sigma}$ is the scatter parameter, z is a nonnegative random variable with density function $p(z)$, and $\boldsymbol{\epsilon}$ is a standard normal random variable with mean zero and covariance matrix identity. z and $\boldsymbol{\epsilon}$ are independent from each other. The mean and the covariance matrix of $\boldsymbol{x}$ are computed as $\boldsymbol{\mu} + \mathsf{E}[z]\boldsymbol{\xi}$ and $\left(\mathsf{E}[z^2] - \mathsf{E}[z]^2\right)\boldsymbol{\xi}\boldsymbol{\xi}^\top + \mathsf{E}[z]\boldsymbol{\Sigma}$, respectively. The conditional density function of $\boldsymbol{x}$ given z is

$$p(\boldsymbol{x} \mid z; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(z\boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2z} \|\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi}\|^2_{\boldsymbol{\Sigma}^{-1}}\right), \tag{4}$$

where $\boldsymbol{\theta} = \left[\boldsymbol{\mu}^\top, \boldsymbol{\xi}^\top, \mathrm{vec}(\boldsymbol{\Sigma})^\top\right]^\top$ and $\|\boldsymbol{x}\|_{\boldsymbol{A}} = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$. Following the definitions of $\boldsymbol{y}$ in (1), the density function of $\boldsymbol{y}$ is given as follows:

$$p(\boldsymbol{y}; \boldsymbol{\theta}) = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x}; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^\infty p(\boldsymbol{x} \mid z; \boldsymbol{\theta}) p(z) \mathrm{d}z \mathrm{d}\boldsymbol{x}, \tag{5}$$

where $\mathcal{Q}^{-1}(\boldsymbol{y})$ maps the quantized data $\boldsymbol{y}$ into a hyper-rectangle whose projection on each dimension i is an interval corresponding to $[\mathcal{Q}(\boldsymbol{y})]_i$. Given n independent and identically distributed samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, the ML estimation problem is given by

$$\max_{\boldsymbol{\theta}} \quad L(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log p(\boldsymbol{y}_t; \boldsymbol{\theta}). \tag{6}$$

When $d = 1$ and z is deterministic, i.e., $\boldsymbol{x}$ is a univariate normal random variable, and one-bit quantization is applied, the estimation problem admits a closed-form solution (Ribeiro & Giannakis, 2006b). In contrast, as long as $d \geq 2$, the density function in (5) involves multiple integrals, making the optimization problem substantially more challenging.

3 THE ECM ALGORITHM

The optimization problem in (6) is challenging due to the multiple integrals in the density function (5). Nevertheless, the ECM framework can be applied by leveraging the models in (1) and (3), where $\boldsymbol{x}$ and z can be naturally treated as latent variables. In the following, we introduce an ECM procedure for (6).

3.1 E-STEP

In the E-step, we derive a surrogate function for the log-likelihood function $L(\boldsymbol{\theta})$. Based on (1), the observed signal $\boldsymbol{y}$ is conditioned on the hidden variable $\boldsymbol{x}$. Hence, given $[\boldsymbol{y}_1 \dots, \boldsymbol{y}_n]^\top$ and the corresponding hidden variables $[\boldsymbol{x}_1 \dots, \boldsymbol{x}_n]^\top$, we have[1]

$$L(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log \int_{\mathbb{R}} p(\boldsymbol{y}_t \mid \boldsymbol{x}_t; \boldsymbol{\theta}) p(\boldsymbol{x}_t; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}_t \geq \sum_{t=1}^{n} \mathsf{E}_{\boldsymbol{x}_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}} \log p(\boldsymbol{y}_t, \boldsymbol{x}_t; \boldsymbol{\theta}) + \text{const.}, \tag{7}$$

where the inequality is based on Jensen's inequality and const. is some constant term independent of $\boldsymbol{\theta}$. Under the models specified in (1), the joint density function is given by

$$p(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{\theta}) = p(\boldsymbol{x}_t; \boldsymbol{\theta}) \delta\left(\mathcal{Q}\left(\boldsymbol{x}_t\right) - \boldsymbol{y}_t\right), \tag{8}$$

where $\delta(\cdot)$ denotes the multivariate Dirac delta function, defined as $\delta\left(\boldsymbol{x}_t\right) = \begin{cases} 0, & \text{if } \boldsymbol{x}_t \neq \boldsymbol{0} \\ +\infty, & \text{if } \boldsymbol{x}_t = \boldsymbol{0} \end{cases}$.

Since the term $\delta\left(\mathcal{Q}\left(\boldsymbol{x}_t\right) - \boldsymbol{y}_t\right)$ is independent of $\boldsymbol{\theta}$, we further obtain

$$L(\boldsymbol{\theta}) \geq \sum_{t=1}^{n} \mathsf{E}_{\boldsymbol{x}_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}} \log p(\boldsymbol{x}_t; \boldsymbol{\theta}) + \text{const.} = \sum_{t=1}^{n} \mathsf{E}_{\boldsymbol{x}_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}} \log \int_{0}^{\infty} p(\boldsymbol{x}_t, z_t; \boldsymbol{\theta}) \mathrm{d}z_t + \text{const.}$$

Regarding z as a hidden variable, applying Jensen's inequality leads to

$$L(\boldsymbol{\theta}) \geq \sum_{t=1}^{n} \mathsf{E}_{\boldsymbol{x}_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}} \left[\mathsf{E}_{z_t|\boldsymbol{x}_t;\underline{\boldsymbol{\theta}}} \log p(\boldsymbol{x}_t, z_t; \boldsymbol{\theta})\right] + \text{const.} \triangleq S(\boldsymbol{\theta}; \underline{\boldsymbol{\theta}}). \tag{9}$$

Since $z \mid \boldsymbol{x}$ is independent of $\boldsymbol{x} \mid \boldsymbol{y}$, we have $p(\boldsymbol{x} \mid \boldsymbol{y}; \boldsymbol{\theta}) p(z \mid \boldsymbol{x}; \boldsymbol{\theta}) = p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta})$. Given the $p(\boldsymbol{x} \mid z; \boldsymbol{\theta})$ in (4), the surrogate function $S(\boldsymbol{\theta}; \underline{\boldsymbol{\theta}})$ can be further expressed as follows:

$$\begin{aligned} S(\boldsymbol{\theta}; \underline{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \Bigg[& \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}} \left[\log p(z_t)\right] - \frac{1}{2} \log \det \boldsymbol{\Sigma} \\ & - \frac{1}{2} \mathrm{Tr}\left(\left(\boldsymbol{U}_t - 2\boldsymbol{v}_t \boldsymbol{\mu}^\top + \iota_t \boldsymbol{\mu}\boldsymbol{\mu}^\top\right) - 2(\boldsymbol{w}_t - \boldsymbol{\mu})\boldsymbol{\xi}^\top + \zeta_t \boldsymbol{\xi}\boldsymbol{\xi}^\top\right) \boldsymbol{\Sigma}^{-1} \Bigg] + \text{const.}, \end{aligned} \tag{10}$$

where

$$\begin{aligned} \boldsymbol{U}_t &= \mathsf{E}_{\boldsymbol{x}_t, z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z^{-1}\boldsymbol{x}\boldsymbol{x}^\top], \quad & \boldsymbol{v}_t &= \mathsf{E}_{\boldsymbol{x}_t, z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z_t^{-1}\boldsymbol{x}_t], \quad & \boldsymbol{w}_t &= \mathsf{E}_{\boldsymbol{x}_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[\boldsymbol{x}_t], \\ \iota_t &= \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z_t^{-1}], & \zeta_t &= \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z_t]. \end{aligned}$$

The details for computing these expectations are given in Appendix A. In the term $\mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}\left[\log p(z_t)\right]$, some scalar parameters (such as the shape parameter ν in Student's t distribution) are contained. In the practical implementation (Galarza et al., 2021), they are typically treated as given or estimated through the one-dimensional search.

3.2 CM-STEP

Based on the surrogate function (10), in the CM-step, we solve the following optimization problem:

$$\max_{\boldsymbol{\theta}} \; S(\boldsymbol{\theta}; \underline{\boldsymbol{\theta}}), \tag{11}$$

where parameters $\boldsymbol{\mu}$, $\boldsymbol{\xi}$, and $\boldsymbol{\Sigma}$ can be solved with closed-form solutions:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{\sum_{t=1}^{n}(\boldsymbol{v}_t - \boldsymbol{\xi})}{\sum_{t=1}^{n} \iota_t}, \qquad \boldsymbol{\xi} = \frac{\sum_{t=1}^{n}(\boldsymbol{w}_t - \boldsymbol{\mu})}{\sum_{t=1}^{n} \zeta_t}, \\ \boldsymbol{\Sigma} &= \frac{1}{n} \sum_{t=1}^{n} \left(\left(\boldsymbol{U}_t - 2\boldsymbol{v}_t \boldsymbol{\mu}^\top + \iota_t \boldsymbol{\mu}\boldsymbol{\mu}^\top\right) - 2(\boldsymbol{w}_t - \boldsymbol{\mu})\boldsymbol{\xi}^\top + \zeta_t \boldsymbol{\xi}\boldsymbol{\xi}^\top\right). \end{aligned} \tag{12}$$

In the context of quantization model estimation, a prototypical scenario is the one-bit Gaussian case (i.e., $e = 1$ and z is a constant). However, this model suffers from an inherent identifiability issue: for each dimension $i = 1, \dots, d$, only the ratio $\frac{\mu_i}{\sigma_i^2}$ can be determined, rather than the individual

[1] Throughout this paper, underlined variables denote those whose values are given as constants.

parameters. Hence, the estimated values of μ_i and σ_i^2 differ from their desirable values by an arbitrary scaling factor. To resolve this, existing estimation methods in the one-bit Gaussian setting typically fix one parameter (either the location vector or the scatter matrix) to identify the other. Moreover, the same ambiguity persists under one-bit quantization whenever $\boldsymbol{x}$ follows any elliptical distribution.

4 Convergence Analysis

In this section, we analyze the convergence properties of the proposed ECM algorithm. The ECM algorithm is guaranteed to converge to a specific stationary point given an initial point (McLachlan & Krishnan, 2008). Define $\boldsymbol{\theta}^\star = \left[\boldsymbol{\mu}^{\star\top}, \boldsymbol{\xi}^{\star\top}, \mathrm{vec}\left(\boldsymbol{\Sigma}^\star\right)^\top\right]^\top$ as a stationary point of the optimization problem in (6). We first establish the convergence of each parameter individually, while keeping the other two fixed, i.e., the convergence property of the conditional maximization step.

Proposition 1. *Denote the sequence generated by the ECM algorithm as* $\left\{\boldsymbol{\theta}^{(k)}\right\}$. *For any* $k \geq 0$, *we have*

$$\begin{aligned}
\|\boldsymbol{\mu}^{(k+1)} - \boldsymbol{\mu}^\star\|_2 &\leq c_\mu \|\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}^\star\|_2, \\
\|\boldsymbol{\xi}^{(k+1)} - \boldsymbol{\xi}^\star\|_2 &\leq c_\xi \|\boldsymbol{\xi}^{(k)} - \boldsymbol{\xi}^\star\|_2, \\
\|\boldsymbol{\Sigma}^{(k+1)} - \boldsymbol{\Sigma}^\star\|_{\mathsf{F}} &\leq c_\Sigma \|\boldsymbol{\Sigma}^{(k)} - \boldsymbol{\Sigma}^\star\|_{\mathsf{F}},
\end{aligned}$$

where

$$\begin{aligned}
c_\mu &= \max_{\boldsymbol{\theta}} \left\| \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1} \mathsf{Cov}_{\boldsymbol{x}_t, z_t | \boldsymbol{y}_t; \boldsymbol{\theta}} \left[z_t^{-1} \boldsymbol{x}_t\right]}{\sum_{t=1}^n \mathsf{E}_{z | \boldsymbol{y}_t; \boldsymbol{\theta}}[z_t^{-1}]} \right\|_2 \in (0,1), \\
c_\xi &= \max_{\boldsymbol{\theta}} \left\| \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1} \mathsf{Cov}_{\boldsymbol{x}_t, z_t | \boldsymbol{y}_t; \boldsymbol{\theta}} \left[\boldsymbol{x}_t - z_t \boldsymbol{\xi}\right]}{\sum_{t=1}^n \mathsf{E}_{z_t | \boldsymbol{y}_t; \boldsymbol{\theta}}[z_t]} \right\|_2 \in (0,1), \\
c_\Sigma &= \max_{\boldsymbol{\theta}} \left\| \frac{1}{2n} \sum_{t=1}^n \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathsf{Cov}_{\boldsymbol{x}_t, z_t | \boldsymbol{y}_t; \boldsymbol{\theta}} \left[z_t^{-1} \mathrm{vec}\left(\boldsymbol{x}_t \boldsymbol{x}_t^\top\right)\right] \right\|_2 \in (0,1).
\end{aligned}$$

Based on the result in Proposition 1 and Meng & Rubin (1994), we establish that the proposed ECM algorithm converges globally at a linear rate, as detailed in the following result.

Theorem 2. *Denote the sequence generated by the ECM algorithm as* $\left\{\boldsymbol{\theta}^{(k)}\right\}$. *For any* $k \geq 0$, *we have*

$$\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^\star\|_2 \leq c^k \|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^\star\|_2, \tag{13}$$

where $c = \max\{c_\mu, c_\xi, c_\Sigma\} \in (0,1)$.

5 Extended Applications

In the previous sections, an algorithm to the ML estimation problem was proposed for modeling quantized data by a normal mean-variance mixture model. Modeling data with a probabilistic model and then performing ML estimation is a common practice in the field of machine learning. Hence, the proposed algorithm can be readily extended to typical machine learning applications from quantized data. In this section, two applications of the proposed model and estimation methods to quantized matrix completion and quantized compressive sensing will be introduced, and experimental validations will be presented in a subsequent section.

5.1 Quantized Matrix Completion

The goal of the low-rank matrix completion problem (Candes & Recht, 2012) is to recover an unknown low-rank matrix $\boldsymbol{M} \in \mathbb{R}^{d_1 \times d_2}$ from an observed, yet incomplete, matrix. Let $\boldsymbol{X}$ denote a matrix whose entries are drawn from a normal mean–variance mixture distribution. We use μ_{ij} and x_{ij} to represent the (i,j)-th entries of $\boldsymbol{M}$ and $\boldsymbol{X}$, respectively. Based on the normal mean-variance model, the relationship between μ_{ij} and x_{ij} is given by

$$x_{ij} = \mu_{ij} + z\xi + z^{1/2}\sigma\epsilon, \tag{14}$$

where μ_{ij} serves as the location parameter of x_{ij}. In the quantization scenario (Davenport et al., 2014), however, the matrix $\boldsymbol{X}$ is not directly accessible; instead, we observe its quantized counterpart $\boldsymbol{Y} = \mathcal{Q}(\boldsymbol{X})$. Moreover, the entire matrix $\boldsymbol{Y}$ is not available for observation. Let $\boldsymbol{\Omega}$ denote the index set of the observed entries; specifically, if $(i, j) \in \boldsymbol{\Omega}$, then Y_{ij} is observed, otherwise Y_{ij} is missing. Our goal is to recover $\boldsymbol{M}$ from incomplete $\boldsymbol{Y}$. The study of matrix completion was popularized following the Netflix Million Dollar Challenge, which posed the task: accurately predicting the values of those entries with a user–item matrix in which entries represent item ratings.

To recover $\boldsymbol{M}$ from quantized and corrupted observations, a seminal study introduced an ML framework for one-bit low-rank matrix completion (Davenport et al., 2014). Building on this work, random dithering was incorporated into the quantization function to improve recovery performance (Eamaz et al., 2024). Both approaches employed the nuclear norm to relax the low-rank constraint and used projected gradient descent for optimization. An alternative strategy reformulated the low-rank constraint via matrix factorization, followed by projected gradient descent (Bhaskar & Javanmard, 2015). Similarly, factorization combined with a majorization–minimization method was used to derive a surrogate objective, which was solved via the Gauss–Newton method (Liu et al., 2025). The framework was further extended from one-bit to multi-bit quantization using low-rank factorization together with projected gradient descent (Bhaskar, 2016).

5.2 Quantized Compressive Sensing

The one-bit compressive sensing (Boufounos & Baraniuk (2008b)) aims to estimate a sparse signal lying in a known measurement subspace based on observed quantized data (Jacques et al., 2013; Chen et al., 2024). Existing methods for this problem generally fall into three categories. The first two primarily focus on 1-bit quantization without incorporating additive noise into the original measurements (Li et al., 2018). The first category, termed regularizer-class algorithms (Laska et al., 2011), introduces additional regularization terms to the classical compressive sensing recovery problem to enforce consistency between the sparse signal and the measurements. The second category, known as penalty-class algorithms (Yan et al., 2012), models sign flips between quantized and recovered measurements as penalty terms. The third category assumes the presence of additive noise in the original measurements (Zymnis et al., 2009; Knudson et al., 2016; Shao et al., 2024). Our proposed model is developed as a further extension of the third approach. To address potential outliers in the additive noise and enhance the robustness of the model, we employ a normal mean–variance mixture model for noise modeling. Denote $\boldsymbol{\Phi} \in \mathbb{R}^{d_1 \times d_2}$ as the known measurement matrix, and $\boldsymbol{\vartheta} \in \mathbb{R}^{d_2}$ as the sparse signal to be estimated. Based on the normal mean-variance model, the quantized compressed sensing model is given by

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{x}), \quad \boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\vartheta} + z\boldsymbol{\xi} + z^{1/2}\sigma\boldsymbol{\epsilon}, \tag{15}$$

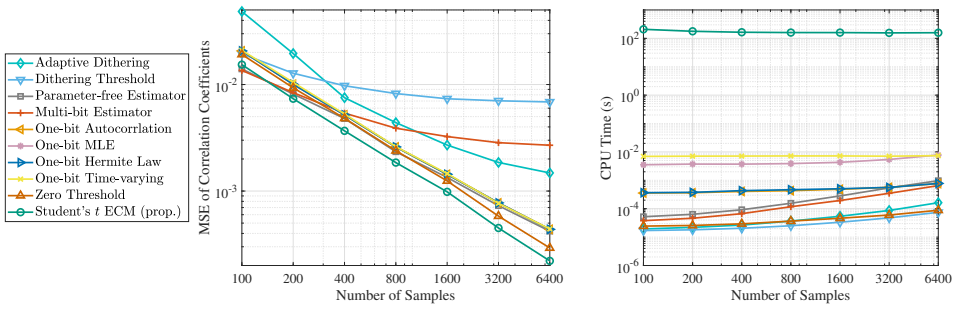
where $\boldsymbol{\Phi}\boldsymbol{\vartheta}$ serves as the location parameter of $\boldsymbol{x}$.

6 Experiments

6.1 Quantized Covariance/Correlation Estimation

In this section, we address the problem of one-bit quantized covariance estimation. Existing approaches can be broadly categorized as follows: non-dithered methods, which include the arcsine law method with a zero threshold (Van Vleck & Middleton, 1966) (Zero Threshold), and methods utilizing a non-zero threshold, such as one-bit autocorrelation estimation (Liu & Lin, 2021) (One-bit Autocorrelation), the one-bit Hermite law (Liu & Chou, 2025) (One-bit Hermite Law), and the covariance matrix recovery method (Xiao et al., 2023) (One-bit MLE). Another category is dithered methods, where the dithering signal follows either a uniform distribution (Dirksen et al., 2022) (Dithering Threshold) or a Gaussian distribution (Eamaz et al., 2022) (One-bit Time-varying); the dithering threshold can also be adaptive based on the data (Dirksen & Maly, 2024) (Adaptive Dithering). For the multi-bit problem, there are the multi-bit covariance estimator (Multi-bit Estimator) and the multi-bit parameter-free estimator (Parameter-free Estimator) (Chen & Ng, 2025).

We begin by comparing the estimation accuracy of our method against existing approaches since the previous work could only estimate the correlation matrix of one-bit data with zero mean. We generated samples from a Student's t distribution with $\nu = 3$ degrees of freedom (Figure 2). We



(a) MSE versus number of samples

(b) CPU time versus number of samples

Figure 1: Algorithm performance comparison for the Student's t distribution

estimated the correlation matrix using the following methods: the Student's t ECM, the Gaussian ECM, the Parameter-free Estimator, and the Multi-bit Estimator. We also compared the results against those obtained by applying the EM algorithm to perform MLE on the unquantized data (Unquantized MLE). The proposed Student's t MLE and Gaussian MLE show that as the number of bits increases, the MSE of estimating correlation will decrease and approach the result of the MLE without quantization. However, as the number of bits increases, the time required by the proposed algorithms will also increase.

Figures 3 and 4 present a comparative analysis of the convergence rates of the algorithm across four statistical distributions: Gaussian, Student's t, GHST, and GH. The results are organized into four subfigures (a–d), each corresponding to one distribution. The convergence is measured by the distance to the optimum, defined as $\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^\star\|_2$ for vectors $\boldsymbol{x}$ and $\boldsymbol{x}^\star$, and as $\|\boldsymbol{X}^{(k)} - \boldsymbol{X}^\star\|_{\mathsf{F}}$ for matrices $\boldsymbol{X}$ and $\boldsymbol{X}^\star$. All proposed methods achieve linear convergence rates. The numerical values of these rates, which vary per iteration, are detailed in Figure 4. Finally, Table 1 reports experimental results where samples are randomly generated from the GH distribution with parameters $\lambda = -0.5$, $\delta = 2$, and $\gamma = 2$, with $e = 1$ and threshold $\tau = 2$. The MSE of the correlation matrix estimated via the GH method is consistently lower than that obtained by alternative approaches.

Table 1: Correlation MSE comparison for different methods under the synthetic data

Method	10^2 samples	10^3 samples	10^4 samples
Gaussian ECM	4.7960×10^{-2}	7.5326×10^{-3}	4.9231×10^{-3}
Student's t ECM	5.8602×10^{-2}	9.5297×10^{-3}	6.6826×10^{-3}
GHST ECM	1.5503×10^{-1}	1.0318×10^{-1}	9.5506×10^{-2}
GH ECM	$\mathbf{4.5377 \times 10^{-2}}$	$\mathbf{7.3532 \times 10^{-4}}$	$\mathbf{9.0011 \times 10^{-4}}$

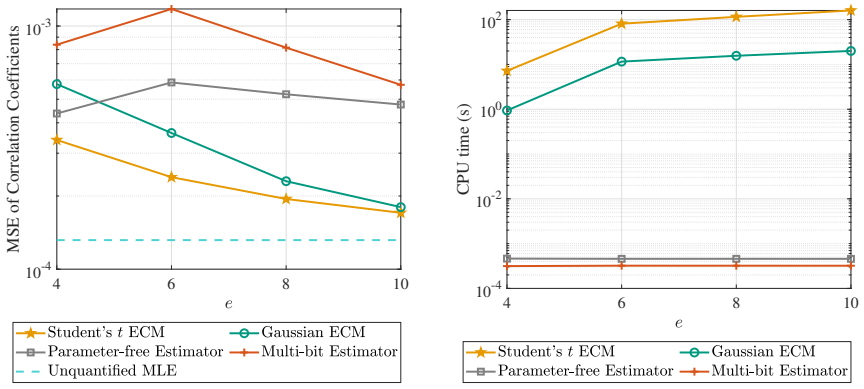
6.2 Quantized Matrix Completion

Given (14) and $\boldsymbol{Y} = \mathcal{Q}(\boldsymbol{X})$, the ML estimation problem for $\boldsymbol{M}$ is given by

$$\begin{aligned} \max_{\boldsymbol{M}} \quad & \sum_{(i,j)\in\boldsymbol{\Omega}} \log p\left(y_{ij} \mid \mu_{ij}\right) \\ \text{s.t.} \quad & \operatorname{rank}(\boldsymbol{M}) \leq r. \end{aligned} \tag{16}$$

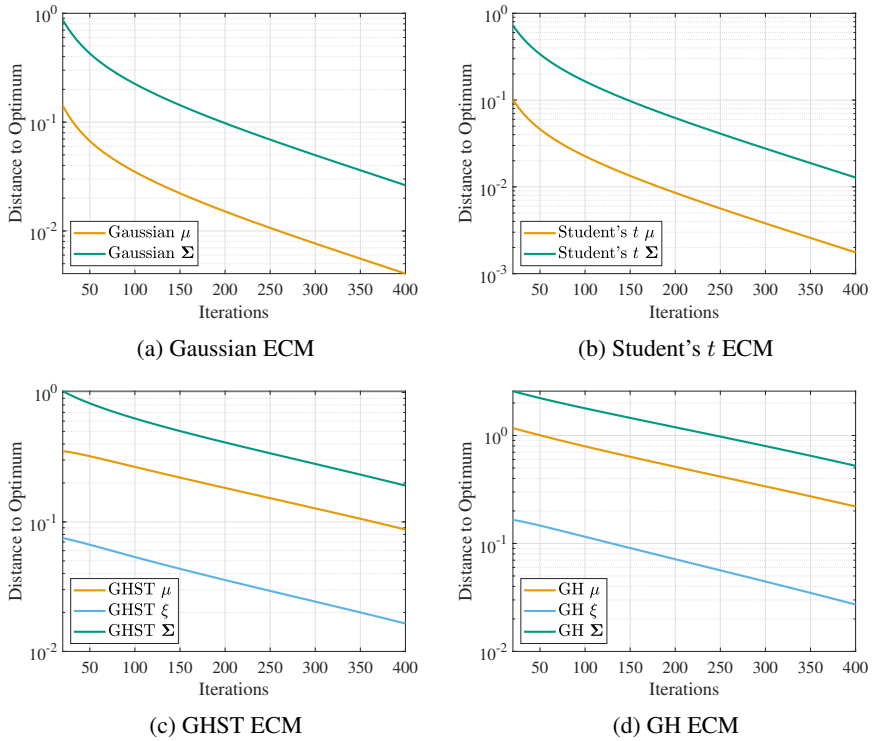
By applying a low-rank factorization to the matrix $\boldsymbol{M}$, we express it as $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{B}^\top$, where $\boldsymbol{A} \in \mathbb{R}^{d_1 \times r}$ and $\boldsymbol{B} \in \mathbb{R}^{d_2 \times r}$. Through the E-step in our proposed ECM algorithm, we have the surrogate function of (16) as

$$\sum_{(i,j)\in\boldsymbol{\Omega}} \left(\boldsymbol{a}_i \boldsymbol{b}_j^\top - e_{ij}\right)^2, \tag{17}$$



(a) MSE versus e (b) CPU time versus e

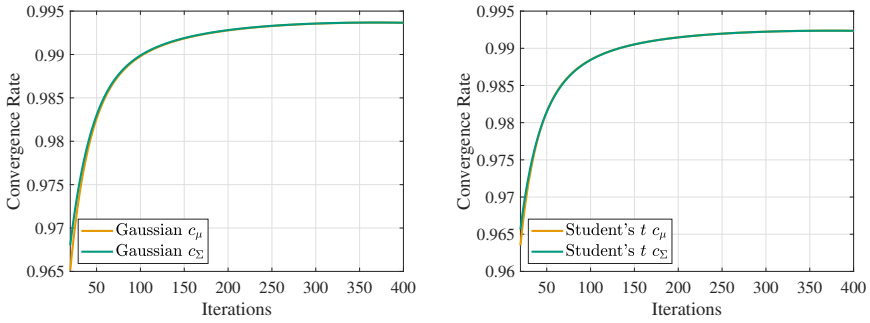
Figure 2: Algorithms performance comparison versus the number of bits



(a) Gaussian ECM (b) Student's t ECM

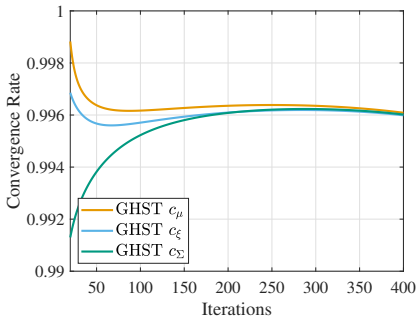
(c) GHST ECM (d) GH ECM

Figure 3: Convergence rate versus iterations

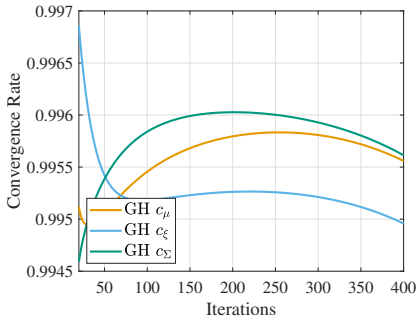


(a) Gaussian ECM

(b) Student's t ECM



(c) GHST ECM



(d) GH ECM

Figure 4: Convergence rate versus iterations

where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are i-th row of the matrix $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, and $e_{ij} = \frac{v_{ij}-\xi}{\iota_{ij}}$. The details of obtaining the surrogate (17) are given in the Appendix D. Then we can use the ECM algorithm to alternatively update $\boldsymbol{A}$ and $\boldsymbol{B}$.

For the experimental design, the MovieLens 1M dataset (Harper & Konstan, 2015) is employed, which contains 1000000 movie ratings provided by 6040 users for 3952 movies, with each rating ranging from 1 to 5. All ratings are randomly partitioned into training and test sets, accounting for 40% and 60% of the data, respectively. The training set is used to predict the completed rating matrix, and the entries overlapping between the completed rating matrix and the test set are then compared to evaluate prediction accuracy and root mean square error (RMSE), which are defined in Appendix E.

Existing methods encompass classic machine learning approaches for matrix completion, including the singular value decomposition (SVD (Sarwar et al., 2000)) model, the ℓ_2-regularized matrix factorization (ℓ_2-regularized (Paterek, 2007)) model, and the nuclear norm minimization (nuclear norm (Cai et al., 2010)) model. Furthermore, several approaches are established upon random variable model assumptions. These include: directly modeling the Gaussian without quantization (Gaussian (Candes & Plan, 2010)); 1-bit quantization with a Gaussian distribution (1-bit Gaussian (Davenport et al., 2014)); and multi-bit quantization under a Gaussian assumption (multi-bit Gaussian (Bhaskar, 2016)). Our proposed method is regarded as an extension of the multi-bit Gaussian approach by extending the Gaussian assumption to a normal mean-variance mixture model, comprising Student's t, GHST, and GH distributions.

The results are demonstrated in Table 2. As shown in the table, the proposed method consistently achieves superior performance in terms of both accuracy and RMSE. Among the evaluated models, the approach based on the GH distribution attains the best results, attributable to its flexibility and robustness in modeling skewness and heavy tails. Details regarding the experimental parameter settings are provided in the Appendix E.

Table 2: Accuracy and RMSE comparisons of matrix completion on MovieLens 1M

Method	Accuracy	RMSE	Time (s)
SVD	0.4261	0.9148	364.36
L2-regularization	0.4388	0.9355	165.08
Nuclear Norm	0.3802	1.1662	867.39
Gaussian	0.4363	0.9486	25.26
1-bit Gaussian	0.4217	0.9659	110.44
Multi-bit Gaussian	0.4453	0.9151	90.84
Multi-bit Student's t ECM (prop.)	0.4464	0.9091	232.16
Multi-bit GHST ECM (prop.)	0.4495	0.9076	252.23
Multi-bit GH ECM (prop.)	**0.4510**	**0.8892**	312.32

6.3 Quantized Compressive Sensing

Given (15) and the ℓ_1 regularization model from Zymnis et al. (2009), we have the optimization problem

$$\max_{\boldsymbol{\vartheta}} \quad \log p\left(\boldsymbol{y} \mid \boldsymbol{\vartheta}\right) + \eta\|\boldsymbol{\vartheta}\|_1. \tag{18}$$

With the E-step in our proposed ECM algorithm, we have the surrogate function of (18) as

$$\frac{1}{2}\left\|\boldsymbol{\Phi}\boldsymbol{\vartheta} - \boldsymbol{e}\right\|_2^2 + \eta\|\boldsymbol{\vartheta}\|_1,$$

where the details of deriving the surrogate function are given in the Appendix D. We can use the FISTA algorithm (Beck & Teboulle, 2009) to solve the surrogate.

Performance comparisons are conducted using synthetic data, where the measurement dimension is denoted as $d_1 = 3000$ and the 30-sparse signal dimension as $d_2 = 1000$. To evaluate algorithmic robustness, an additive noise term exhibiting both skewed and heavy-tailed characteristics is introduced to the original measurements. The existing methods included in our comparison consist of restricted step shrinkage (RSS) (Laska et al., 2011), adaptive outlier pursuit (AOP) (Yan et al., 2012), and 1-bit Gaussian MLE (Zymnis et al., 2009). For each signal-to-noise ratio (SNR) level, experiments are repeated 10 times, and the average cosine similarity (Cos Sim) and computational time are reported. The results are summarized in the Table 3, which demonstrates that the 1-bit GH ECM algorithm achieves the largest average cosine similarity. The definitions of SNR and cosine similarity are given in Appendix E.

Table 3: Cosine similarity comparisons of 1-bit compressive sensing under different SNR levels.

Method	SNR = 0 dB		SNR = -5 dB		SNR = -10 dB	
	Cos Sim	Time (s)	Cos Sim	Time (s)	Cos Sim	Time (s)
RSS	0.7872	0.2639	0.6598	0.2651	0.3027	0.2546
AOP	0.7052	0.2689	0.4041	0.2951	0.2209	0.2738
1-bit Gaussian	0.8220	2.7704	0.5487	3.0875	0.2896	2.7202
1-bit Student's t ECM (prop.)	0.8396	2.8573	0.5678	3.1643	0.3031	3.3174
1-bit GHST ECM (prop.)	0.8893	2.7435	0.7701	3.4204	0.5237	4.0281
1-bit GH ECM (prop.)	**0.8915**	2.7982	**0.7717**	3.3562	**0.5407**	3.9535

7 Conclusion and Discussion

In this paper, we propose an ECM-based algorithm for parameter estimation in quantized models. The method exhibits excellent scalability, handling problems from one-bit to multi-bit quantization and accommodating distributions ranging from Gaussian to the broader class of normal mean–variance mixtures. Experiments show that our approach yields more accurate estimates than existing methods in certain cases (e.g., the one-bit Gaussian setting), while maintaining high accuracy across a wide range of extended scenarios. Furthermore, it demonstrates strong potential in the typical machine learning tasks including quantized matrix completion and quantized compressive sensing.

ETHICS STATEMENT

All authors of this paper have read and agree to comply with the ICLR Code of Ethics (https://iclr.cc/public/CodeOfEthics). This work does not involve human subjects, sensitive data, or applications that could directly cause harm. The datasets employed in our experiments are publicly available and do not contain personally identifiable information. We have taken measures to ensure that our methodology does not introduce or exacerbate unfair bias, and potential limitations as well as societal impacts are discussed in the main text. There are no conflicts of interest or external sponsorships that could influence the results or their interpretation.

REPRODUCIBILITY STATEMENT

We are committed to ensuring the reproducibility of our results. All experimental details, including model architectures, hyperparameters, training procedures, and evaluation metrics, are described in Section 6 and Appendix E. The datasets used are publicly accessible, and we include the version information in Section 6.

REFERENCES

Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-gaussian measurements. *Linear Algebra and its Applications*, 441:222–239, 2014. 1

Aria Ameri, Arindam Bose, Jian Li, and Mojtaba Soltanalian. One-bit radar processing with time-varying sampling thresholds. *IEEE Transactions on Signal Processing*, 67(20):5297–5308, 2019. 1

Ofer Bar-Shalom and Anthony J Weiss. DOA estimation using one-bit quantized measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):868–884, 2002a. 1

Ofer Bar-Shalom and Anthony J Weiss. DOA estimation using one-bit quantized measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):868–884, 2002b. 1

Ole Barndorff-Nielsen, John Kent, and Michael Sørensen. Normal variance-mean mixtures and z distributions. *International Statistical Review*, 50(2):145, 1982. 3

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 10

Sonia A. Bhaskar. Probabilistic low-rank matrix completion from quantized measurements. *Journal of Machine Learning Research*, 17(60):1–34, 2016. 1, 6, 9

Sonia A Bhaskar and Adel Javanmard. 1-bit matrix completion under exact low-rank constraint. In *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6. IEEE, 2015. 6

Giulio Bottegal and Johan A. K. Suykens. Probabilistic matrix factorization from quantized measurements. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 270–277, 2017. 1

Petros T. Boufounos and Richard G. Baraniuk. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pp. 16–21, 2008a. 1

Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pp. 16–21. IEEE, 2008b. 6

Herm Jan Brascamp and Elliott H Lieb. On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of functional analysis*, 22(4):366–389, 1976. 20

Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. 9

Tony Cai and Wen-Xin Zhou. A max-norm constrained minimization approach to 1-bit matrix completion. *Journal of Machine Learning Research*, 14(1):3619–3647, 2013. 1

Emmanuel Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6): 925–936, 2010. 9

Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012. 5

George Casella and Roger Berger. *Statistical inference*. CRC press, 2024. 18

Junren Chen and Michael K Ng. A parameter-free two-bit covariance estimator with improved operator norm error rate. *Applied and Computational Harmonic Analysis*, pp. 101774, 2025. 6

Junren Chen, Yueqi Wang, and Michael K Ng. Quantized low-rank multivariate regression with random dithering. *IEEE Transactions on Signal Processing*, 71:3913–3928, 2023. 1

Junren Chen, Michael K. Ng, and Di Wang. Quantizing heavy-tailed data in statistical estimation: (near) minimax rates, covariate quantization, and uniform recovery. *IEEE Transactions on Information Theory*, 70(3):2003–2038, 2024. 2, 6

Yuejie Chi and Haoyu Fu. Subspace learning from bits. *IEEE Transactions on Signal Processing*, 65(17):4429–4442, 2017. 1

Junil Choi, Jianhua Mo, and Robert W. Heath. Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs. *IEEE Transactions on Communications*, 64(5):2005–2018, 2016. 1

Onkar Dabeer and Elias Masry. Multivariate signal parameter estimation under dependent noise from 1-bit dithered quantized data. *IEEE Transactions on Information Theory*, 54(4):1637–1654, 2008. 1, 2

Mark A. Davenport, Yaniv Plan, Ewout Van Den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference*, 3(3):189–223, 2014. 1, 6, 9

Sjoerd Dirksen and Johannes Maly. Tuning-Free One-Bit Covariance Estimation Using Data-Driven Dithering. *IEEE Transactions on Information Theory*, 70(7):5228–5247, July 2024. 6

Sjoerd Dirksen, Johannes Maly, and Holger Rauhut. Covariance estimation under one-bit quantization. *The Annals of Statistics*, 50(6):3538–3562, 2022. 1, 2, 6

Sjoerd Dirksen, Weilin Li, and Johannes Maly. Subspace estimation under coarse quantization. In *2025 International Conference on Sampling Theory and Applications (SampTA)*, pp. 1–5. IEEE, 2025. 1

Arian Eamaz, Farhang Yeganegi, and Mojtaba Soltanalian. Covariance recovery for one-bit sampled non-stationary signals with time-varying sampling thresholds. *IEEE Transactions on Signal Processing*, 70:5222–5236, 2022. 1, 2, 6

Arian Eamaz, Farhang Yeganegi, and Mojtaba Soltanalian. Covariance recovery for one-bit sampled stationary signals with time-varying sampling thresholds. *Signal Processing*, 206:108899, 2023. 1, 2

Arian Eamaz, Farhang Yeganegi, and Mojtaba Soltanalian. Matrix completion from one-bit dither samples. *IEEE Transactions on Signal Processing*, pp. 1–14, 2024. 6

Jun Fang and Hongbin Li. Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood. *IEEE Transactions on Signal Processing*, 56(10):5246–5257, 2008. 1, 2

Christian E. Galarza, Tsung-I Lin, Wan-Lun Wang, and Víctor H. Lachos. On moments of folded and truncated multivariate student-t distributions based on recurrence relations. *Metrika*, 84(6): 825–850, 2021. 4

Pengzhi Gao, Ren Wang, Meng Wang, and Joe H. Chow. Low-rank matrix recovery from noisy, quantized, and erroneous measurements. *IEEE Transactions on Signal Processing*, 66(11):2918–2932, 2018. 1

Robert M. Gray and Thomas G. Stockham. Dithered quantizers. *IEEE Transactions on Information Theory*, 39(3):805–812, 1993. 1

László Gyorfi and Marten Wegkamp. Quantization for nonparametric regression. *IEEE Transactions on Information Theory*, 54(2):867–874, 2008. 1

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM transactions on interactive intelligent systems*, 5(4):1–19, 2015. 9

Hsiu J Ho, Tsung-I Lin, Hsuan-Yu Chen, and Wan-Lun Wang. Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*, 142(1):25–40, 2012. 16

Laurent Jacques, Jason N. Laska, Petros T. Boufounos, and Richard G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013. 6

Benzhou Jin, Jiang Zhu, Qihui Wu, Yuhong Zhang, and Zhiwei Xu. One-bit lfmcw radar: Spectrum analysis and target detection. *IEEE Transactions on Aerospace and Electronic Systems*, 56(4): 2732–2750, 2020. 1

Karin Knudson, Rayan Saab, and Rachel Ward. One-bit compressive sensing with norm estimation. *IEEE Transactions on Information Theory*, 62(5):2748–2758, 2016. 6

Jason N Laska, Zaiwen Wen, Wotao Yin, and Richard G Baraniuk. Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements. *IEEE Transactions on Signal Processing*, 59(11):5289–5301, 2011. 6, 10

Paul Lévy. Sur certains processus stochastiques homogènes. *Compositio mathematica*, 7:283–339, 1940. 2

Zhilin Li, Wenbo Xu, Xiaobo Zhang, and Jiaru Lin. A survey on one-bit compressed sensing: Theory and applications. *Frontiers of Computer Science*, 12(2):217–230, 2018. 6

Chun-Lin Liu and Yi-Hung Chou. Approximation and analysis of the one-bit Hermite law. In *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2025. 2, 6

Chun-Lin Liu and Zi-Min Lin. One-bit autocorrelation estimation with non-zero thresholds. In *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4520–4524. IEEE, 2021. 1, 2, 6

Xiaoqian Liu, Xu Han, Eric C. Chi, and Boaz Nadler. A majorization-minimization gauss-newton method for 1-bit matrix completion. *Journal of Computational and Graphical Statistics*, 34(3): 1017–1029, 2025. 6

Xicheng Lu, Wei Liu, and Akram Alomainy. A 1.5-bit quantization scheme and its application to sparse array direction estimation. In *2024 19th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–5, 2024. 1

David Q. Mayne. The use of quantization in regression analysis. *International Journal of Control*, 6(6):573–577, 1967. 1

Geoffrey J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 2008. 5, 18

Xiao-Li Meng and Donald B Rubin. On the global and componentwise rates of convergence of the em algorithm. *Linear Algebra and its Applications*, 199:413–425, 1994. 5, 28

Jianhua Mo, Philip Schniter, and Robert W Heath. Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs. *IEEE Transactions on Signal Processing*, 66(5):1141–1154, 2017. 1

Haralabos C. Papadopoulos, Gregory W Wornell, and Alan V Oppenheim. Sequential signal encoding from noisy measurements using quantizers with dynamic bias control. *IEEE Transactions on Information Theory*, 47(3):978–1002, 2001. 1, 2

Arkadiusz Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of the KDD Cup and Workshop*, pp. 394–401. ACM, 2007. 9

Daniel Plabst, Jawad Munir, Amine Mezghani, and Josef A. Nossek. Efficient non-linear equalization for 1-bit quantized cyclic prefix-free massive MIMO systems. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6, 2018. 1

Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2013. 1

Jiaying Ren and Jian Li. One-bit digital radar. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pp. 1142–1146, 2017. 1

Alejandro Ribeiro and Georgios B Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part II: Unknown probability density function. *IEEE Transactions on Signal Processing*, 54(7):2784–2796, 2006a. 1, 2

Alejandro Ribeiro and Georgios B Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case. *IEEE Transactions on Signal Processing*, 54(3): 1131–1143, 2006b. 1, 2, 3

Lawrence Roberts. Picture coding using pseudo-random noise. *IRE Transactions on Information Theory*, 8(2):145–154, 1962. 1

B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl. Application of dimensionality reduction in recommender systems: A case study. In *WebKDD Workshop at the ACM SIGKKD*. ACM, 2000. 9

Mingjie Shao, Wing-Kin Ma, Junbin Liu, and Zihao Huang. Accelerated and deep expectation maximization for one-bit MIMO-OFDM detection. *IEEE Transactions on Signal Processing*, 72: 1094–1113, 2024. 1, 6

Christoph Stöckle, Jawad Munir, Amine Mezghani, and Josef A. Nossek. Channel estimation in massive MIMO systems using 1-bit quantization. In *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–6, 2016. 1

J.H. Van Vleck and D. Middleton. The spectrum of clipped noise. *Proceedings of the IEEE*, 54(1): 2–19, 1966. 1, 2, 6

Bernard Widrow and István Kollár. *Quantization noise: roundoff error in digital computation, signal processing, control, and communications*. Cambridge University Press, 2008. 1

Yu-Hang Xiao, Lei Huang, David Ramírez, Cheng Qian, and Hing Cheung So. Covariance matrix recovery from one-bit data with non-zero quantization thresholds: Algorithm and performance analysis. *IEEE Transactions on Signal Processing*, 71:4060–4076, 2023. 1, 2, 6

Ming Yan, Yi Yang, and Stanley Osher. Robust 1-bit compressive sensing using adaptive outlier pursuit. *IEEE Transactions on Signal Processing*, 60(7):3868–3875, 2012. 6, 10

Jian Yang, Zihang Song, Han Zhang, and Yue Gao. Compressive spectrum sensing with 1-bit ADCs. *IEEE Transactions on Vehicular Technology*, 2025. 1

Argyrios Zymnis, Stephen Boyd, and Emmanuel Candes. Compressed sensing with quantized measurements. *IEEE Signal Processing Letters*, 17(2):149–152, 2009. 1, 6, 10, 29

LLM USAGE STATEMENT

During the preparation of this manuscript, we employed large language models (LLMs) as general-purpose writing and editing tools. LLMs were utilized to enhance the clarity, grammar, and structure of the text, but they did not contribute to research ideation, experimental design, or scientific content. No LLM was used to generate new ideas, mathematical proofs, or experimental results. All content, including any text suggested by LLMs, was thoroughly reviewed and verified by the authors, who assume full responsibility for the final manuscript.

A EXPECTATION EVALUATIONS ON SPECIFIC DISTRIBUTION ASSUMPTIONS

From Section 3, we have established that when the random variable $\boldsymbol{x}$ in model (3) belongs to the normal mean–variance mixture family, the surrogate function (10) can be derived, along with its closed-form solutions with respect to $\boldsymbol{\mu}$, $\boldsymbol{\xi}$, and $\boldsymbol{\Sigma}$. Nevertheless, these closed-form solutions (12) depend on the expected values $\boldsymbol{U}_t$, $\boldsymbol{v}_t$, $\boldsymbol{w}_t$, ι_t, and ζ_t, which are determined by the distribution of the random variable $\boldsymbol{x}$. Therefore, in the following, we analyze the integrals corresponding to these expected values under various distributions of $\boldsymbol{x}$, and examine the convergence properties and statistical characteristics of the proposed algorithm under specific distributional assumptions. We first consider the fundamental case of the Gaussian distribution (i.e., $z = 1$ and $\boldsymbol{\xi}$). Let $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In this case, the first- and second-order moments are given by Ho et al. (2012) as

$$\mathsf{E}_{\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}] = \boldsymbol{\mu} + p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\,\boldsymbol{\Sigma}\boldsymbol{q}, \tag{19}$$

$$\mathsf{E}_{\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}\boldsymbol{x}^\top\right] = \boldsymbol{\mu}\boldsymbol{\mu}^\top + \boldsymbol{\Sigma} + p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\left(2\left(\boldsymbol{\Sigma}\boldsymbol{q}\right)\boldsymbol{\mu}^\top + \boldsymbol{\Sigma}\left(\boldsymbol{H}+\boldsymbol{D}\right)\boldsymbol{\Sigma}\right). \tag{20}$$

Let the interval $\mathcal{Q}^{-1}(y_i)$ be $[l_i, u_i]$. The i-th elements of vector $\boldsymbol{q}$ is given by

$$q_i = p(x_i = l_i, \boldsymbol{y}_{\setminus i};\boldsymbol{\theta}) - p(x_i = u_i, \boldsymbol{y}_{\setminus i};\boldsymbol{\theta}), \tag{21}$$

The matrix $\boldsymbol{H}$ is a matrix with all diagonal entries being zero and the (i,j)-th off-diagonal element being

$$\boldsymbol{H}_{ij} = h(l_i, l_j) - h(l_i, u_j) - h(u_i, l_j) + h(u_i, u_j), \tag{22}$$

with

$$h(c_i, c_j) = p(x_i = c_i, x_j = c_j, \boldsymbol{y}_{\setminus i,j};\boldsymbol{\theta})).$$

The matrix $\boldsymbol{D}$ is a diagonal matrix with the diagonal entries:

$$\boldsymbol{D}_{ii} = \frac{l_i - \mu_i}{\sigma_i^2} p(x_i = l_i, \boldsymbol{y}_{\setminus i};\boldsymbol{\theta}) - \frac{u_i - \mu_i}{\sigma_i^2} p(x_i = u_i, \boldsymbol{y}_{\setminus i};\boldsymbol{\theta}) - \frac{[\boldsymbol{\Sigma}\boldsymbol{H}]_{ii}}{\sigma_i^2}. \tag{23}$$

Now we consider the general normal mean-variance mixture with $\boldsymbol{x} = \boldsymbol{\mu} + z\boldsymbol{\xi} + (z\boldsymbol{\Sigma})^{\frac{1}{2}}\boldsymbol{\epsilon}$. In this case, the first and second moments satisfy

$$\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}] = \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[\mathsf{E}_{\boldsymbol{x}|z,\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}]\right] \text{ and } \mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}\boldsymbol{x}^\top\right] = \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[\mathsf{E}_{\boldsymbol{x}|z,\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}\boldsymbol{x}^\top]\right].$$

Based on $\boldsymbol{x} \mid z \sim \mathcal{N}(\boldsymbol{\mu} + z\boldsymbol{\xi}, z\boldsymbol{\Sigma})$ and the moments in (19) and (20), we have

$$\mathsf{E}_{\boldsymbol{x}|z,\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}] = \boldsymbol{\mu} + z\boldsymbol{\xi} + p^{-1}(\boldsymbol{y} \mid z;\boldsymbol{\theta}) z\boldsymbol{\Sigma}\boldsymbol{q}_z, \tag{24}$$

$$\begin{aligned}\mathsf{E}_{\boldsymbol{x}|z,\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}\boldsymbol{x}^\top\right] = {} & \boldsymbol{\mu}\boldsymbol{\mu}^\top + z^2\boldsymbol{\xi}\boldsymbol{\xi}^\top + 2z\boldsymbol{\mu}\boldsymbol{\xi}^\top + z\boldsymbol{\Sigma} \\ & + p^{-1}(\boldsymbol{y} \mid z;\boldsymbol{\theta})\left(2z\left(\boldsymbol{\Sigma}\boldsymbol{q}_z\right)\boldsymbol{\mu}^\top + 2z^2\left(\boldsymbol{\Sigma}\boldsymbol{q}_z\right)\boldsymbol{\xi}^\top + z^2\boldsymbol{\Sigma}\left(\boldsymbol{H}_z + \boldsymbol{D}_z\right)\boldsymbol{\Sigma}\right),\end{aligned} \tag{25}$$

where the i-th elements of vector $\boldsymbol{q}_z$ is denoted as

$$q_{z,i} = p(x_i = l_i, \boldsymbol{y}_{\setminus i} \mid z;\boldsymbol{\theta}) - p(x_i = u_i, \boldsymbol{y}_{\setminus i} \mid z;\boldsymbol{\theta}), \tag{26}$$

the matrix $\boldsymbol{H}_z$ is a matrix with all diagonal entries being zero and the (i,j)-th off-diagonal element being

$$\boldsymbol{H}_{z,ij} = h_z(l_i, l_j) - h_z(l_i, u_j) - h_z(u_i, l_j) + h_z(u_i, u_j), \tag{27}$$

with

$$h_z(c_i, c_j) = p(x_i = c_i, x_j = c_j, \boldsymbol{y}_{\setminus i,j} \mid z;\boldsymbol{\theta})),$$

and the matrix $\boldsymbol{D}_z$ is a diagonal matrix with the diagonal entries:

$$\boldsymbol{D}_{z,ii} = \frac{l_i - \mu_i - z\xi_i}{\sigma_i^2} p(x_i = l_i, \boldsymbol{y}_{\setminus i} \mid z; \boldsymbol{\theta}) - \frac{u_i - \mu_i - z\xi_i}{\sigma_i^2} p(x_i = u_i, \boldsymbol{y}_{\setminus i} \mid z; \boldsymbol{\theta}) - \frac{[\boldsymbol{\Sigma}\boldsymbol{H}_z]_{ii}}{\sigma_i^2}. \tag{28}$$

Based on (24) and (25), we can obtain

$$\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}] = \boldsymbol{\mu} + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\,\boldsymbol{\xi} + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[zp^{-1}(\boldsymbol{y} \mid z; \boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{q}_z\right], \tag{29}$$

$$\begin{aligned}\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}\boldsymbol{x}^\top\right] = {} & \boldsymbol{\mu}\boldsymbol{\mu}^\top + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^2\right]\boldsymbol{\xi}\boldsymbol{\xi}^\top + 2\,\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\,\boldsymbol{\mu}\boldsymbol{\xi}^\top + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\,\boldsymbol{\Sigma} \\ & + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[p^{-1}(\boldsymbol{y} \mid z; \boldsymbol{\theta})\left(2z\left(\boldsymbol{\Sigma}\boldsymbol{q}_z\right)\boldsymbol{\mu}^\top + 2z^2\left(\boldsymbol{\Sigma}\boldsymbol{q}_z\right)\boldsymbol{\xi}^\top + z^2\boldsymbol{\Sigma}\left(\boldsymbol{H}_z + \boldsymbol{D}_z\right)\boldsymbol{\Sigma}\right)\right].\end{aligned} \tag{30}$$

We first compute $\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^k\right]$ as follows:

$$\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^k\right] = \int_0^{+\infty} p(z \mid \boldsymbol{y}; \boldsymbol{\theta}) z^k \mathrm{d}z = \int_0^{+\infty} \frac{p(\boldsymbol{y} \mid z; \boldsymbol{\theta})p(z)}{p(\boldsymbol{y}; \boldsymbol{\theta})} z^k \mathrm{d}z.$$

Here, we introduce the size-biased distribution of order k of the positive random variable z, which has the density function $p_k(z) = \frac{z^k p(z)}{\mathsf{E}_z[z^k]}$. Hence, we have

$$\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^k\right] = \mathsf{E}_z[z^k] \frac{\int_0^{+\infty} p(\boldsymbol{y} \mid z; \boldsymbol{\theta}) p_k(z) \mathrm{d}z}{p(\boldsymbol{y}; \boldsymbol{\theta})} = \mathsf{E}_z[z^k] \frac{p_k(\boldsymbol{y}; \boldsymbol{\theta})}{p(\boldsymbol{y}; \boldsymbol{\theta})}, \tag{31}$$

where we denote $p_k(\boldsymbol{y}; \boldsymbol{\theta}) = \int_0^{+\infty} p(\boldsymbol{y} \mid z; \boldsymbol{\theta}) p_k(z) \mathrm{d}z$.

Similarly, we can obtain

$$\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^k p^{-1}(\boldsymbol{y} \mid z; \boldsymbol{\theta})\boldsymbol{q}_z\right] = \mathsf{E}_z[z^k] p^{-1}(\boldsymbol{y}; \boldsymbol{\theta})\boldsymbol{q}_k, \tag{32}$$

where $\boldsymbol{q}_k = \int_0^{+\infty} \boldsymbol{q}_z p_k(z) \mathrm{d}z$ with the i-th element

$$q_{k,i} = p_k(x_i = l_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) - p_k(x_i = u_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}). \tag{33}$$

For the expectation of $z^k \boldsymbol{H}_z$ we have

$$\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^k p^{-1}(\boldsymbol{y} \mid z; \boldsymbol{\theta})\left(\boldsymbol{H}_z + \boldsymbol{D}_z\right)\right] = p^{-1}(\boldsymbol{y}; \boldsymbol{\theta})\left(\boldsymbol{H}_k + \boldsymbol{D}_k\right), \tag{34}$$

with the (i, j) entry of

$$\boldsymbol{H}_{k,ij} = \mathsf{E}_z[z^k]\left(h_k(l_i, l_j) - h_k(l_i, u_j) - h_k(u_i, l_j) + h_k(u_i, u_j)\right), \tag{35}$$

$$h_k(c_i, c_j) = p_k(x_i = c_i, x_j = c_j, \boldsymbol{y}_{\setminus i,j}; \boldsymbol{\theta})), \tag{36}$$

and the matrix $\boldsymbol{D}_k$ is a diagonal matrix with the diagonal entries:

$$\begin{aligned}\boldsymbol{D}_{k,ii} = {} & \mathsf{E}_z[z^{k-1}] \frac{l_i - \mu_i}{\sigma_i^2} p_{k-1}(x_i = l_i, \boldsymbol{y}_{\setminus i} \mid z; \boldsymbol{\theta}) - \mathsf{E}_z[z^{k-1}] \frac{u_i - \mu_i}{\sigma_i^2} p_{k-1}(x_i = u_i, \boldsymbol{y}_{\setminus i} \mid z; \boldsymbol{\theta}) \\ & - \mathsf{E}_z[z^k] \frac{\xi_i}{\sigma_i^2} q_{k,i} - \frac{[\boldsymbol{\Sigma}\boldsymbol{H}_k]_{ii}}{\sigma_i^2}.\end{aligned} \tag{37}$$

Therefore, the expectations in (29) and (30) become

$$\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}[\boldsymbol{x}] = \boldsymbol{\mu} + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\,\boldsymbol{\xi} + \mathsf{E}_z[z]\, p^{-1}(\boldsymbol{y}; \boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{q}_1, \tag{38}$$

$$\begin{aligned}\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}\boldsymbol{x}^\top\right] = {} & \boldsymbol{\mu}\boldsymbol{\mu}^\top + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^2\right]\boldsymbol{\xi}\boldsymbol{\xi}^\top + 2\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\,\boldsymbol{\mu}\boldsymbol{\xi}^\top + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\,\boldsymbol{\Sigma} \\ & + p^{-1}(\boldsymbol{y}; \boldsymbol{\theta})\left(2\mathsf{E}_z[z]\left(\boldsymbol{\Sigma}\boldsymbol{q}_1\right)\boldsymbol{\mu}^\top + 2\mathsf{E}_z\left[z^2\right]\left(\boldsymbol{\Sigma}\boldsymbol{q}_2\right)\boldsymbol{\xi}^\top + \boldsymbol{\Sigma}\left(\boldsymbol{H}_2 + \boldsymbol{D}_2\right)\boldsymbol{\Sigma}\right).\end{aligned} \tag{39}$$

Meanwhile, we can also obtain

$$\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^{-1}\boldsymbol{x}\right] = \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[\mathsf{E}_{\boldsymbol{x}|z,\boldsymbol{y};\boldsymbol{\theta}}[z^{-1}\boldsymbol{x}]\right] = \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^{-1}\right]\boldsymbol{\mu} + \boldsymbol{\xi} + p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{q}, \tag{40}$$

$$\begin{aligned}\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^{-1}\boldsymbol{x}\boldsymbol{x}^\top\right] &= \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[\mathsf{E}_{\boldsymbol{x}|z,\boldsymbol{y};\boldsymbol{\theta}}[z^{-1}\boldsymbol{x}\boldsymbol{x}^\top]\right] \\ &= \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z^{-1}\right]\boldsymbol{\mu}\boldsymbol{\mu}^\top + \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z\right]\boldsymbol{\xi}\boldsymbol{\xi}^\top + 2\boldsymbol{\mu}\boldsymbol{\xi}^\top + \boldsymbol{\Sigma} \\ &\quad + p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\left(2\left(\boldsymbol{\Sigma}\boldsymbol{q}\right)\boldsymbol{\mu}^\top + 2\mathsf{E}_z\left[z\right]\left(\boldsymbol{\Sigma}\boldsymbol{q}_1\right)\boldsymbol{\xi}^\top + \boldsymbol{\Sigma}\left(\boldsymbol{H}_1 + \boldsymbol{D}_1\right)\boldsymbol{\Sigma}\right).\end{aligned} \tag{41}$$

B THE PROOF OF PROPOSITION 1

In this section, we prove the global linear convergence of our proposed ECM algorithm with respect to the three parameters: $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\xi}$. Since the ECM algorithm can always converge to the stationary points (McLachlan & Krishnan, 2008), we define $\boldsymbol{\mu}^\star$, $\boldsymbol{\Sigma}^\star$, and $\boldsymbol{\xi}^\star$ as the stationary points of the optimization problem (6). In the following, we give the convergence analysis of these three parameters with the convergence rate, respectively.

B.1 CONVERGENCE RATE OF LOCATION PARAMETER

Given the update rule of $\boldsymbol{\mu}$ [2] in (12), we have

$$\boldsymbol{\mu} = \frac{\sum_{t=1}^n(\boldsymbol{v}_t - \underline{\boldsymbol{\xi}})}{\sum_{t=1}^n \iota_t} = \frac{\sum_{t=1}^n(\mathsf{E}_{\boldsymbol{x}_t,z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z^{-1}\boldsymbol{x}] - \underline{\boldsymbol{\xi}})}{\sum_{t=1}^n \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z^{-1}]}. \tag{42}$$

Based on (31) and (40), the update rule of $\boldsymbol{\mu}$ in (42) becomes

$$\boldsymbol{\mu} = \underline{\boldsymbol{\mu}} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}{p(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}\right)^{-1}\sum_{t=1}^n p^{-1}(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}\boldsymbol{q}_t \tag{43}$$

Hence, the distance between $\boldsymbol{\mu}$ at the current iteration and $\boldsymbol{\mu}^\star$ is given by

$$\|\boldsymbol{\mu} - \boldsymbol{\mu}^\star\|_2 = \left\|\underline{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}{p(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}\right)^{-1}\sum_{t=1}^n p^{-1}(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}\boldsymbol{q}_t\right\|_2. \tag{44}$$

To further analyze the term on the right side in (44), we first give a result of $\boldsymbol{q}$.

Lemma 3. *When $\boldsymbol{x}$ follows a normal mean-variance mixture and $\boldsymbol{q}$ is defined in (21), we have*

$$\boldsymbol{q} = \nabla_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta}).$$

Proof. Consider the partial derivative of $p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to the i-th element of $\boldsymbol{\mu}$. We have that

$$\nabla_{\mu_i} p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\mu_i}\int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x} = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \nabla_{\mu_i} p(\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x},$$

where the interchange of the integration and differentiation operations in the second equal sign is valid as the Leibniz integral rule (Casella & Berger, 2024). Since the partial derivatives of $p(\boldsymbol{x};\boldsymbol{\theta})$

$$\nabla_{\mu_i} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) = -\nabla_{x_i} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}),$$

we can obtain that

$$\begin{aligned}\nabla_{\mu_i} p(\boldsymbol{y};\boldsymbol{\theta}) &= -\int_0^{+\infty}\int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i})}\int_{l_i}^{u_i} \nabla_{x_i} p(\boldsymbol{x} \mid z;\boldsymbol{\theta})p(z)\mathrm{d}x_i\mathrm{d}\boldsymbol{x}_{\setminus i}\mathrm{d}z \\ &= \int_0^{+\infty}\int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i})} p(z)\left(p(\boldsymbol{x}_{\setminus i}, x_i = l_i \mid z;\boldsymbol{\theta}) - p(\boldsymbol{x}_{\setminus i}, x_i = u_i \mid z;\boldsymbol{\theta})\right)\mathrm{d}\boldsymbol{x}_{\setminus i}\mathrm{d}z \\ &= p(x_i = l_i, \boldsymbol{y}_{\setminus i};\boldsymbol{\theta}) - p(x_i = u_i, \boldsymbol{y}_{\setminus i};\boldsymbol{\theta}),\end{aligned}$$

which is equivalent to the definition of the i-th element of $\boldsymbol{q}$ in (21).

[2] For simplicity, we use $\underline{\boldsymbol{\mu}}$, $\underline{\boldsymbol{\xi}}$, and $\underline{\boldsymbol{\Sigma}}$ to denote the parameters $\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\xi}^{(k)}$, and $\boldsymbol{\Sigma}^{(k)}$ before the k-th update, and use $\boldsymbol{\mu}$, $\boldsymbol{\xi}$, and $\boldsymbol{\Sigma}$ to denote the updated parameters $\boldsymbol{\mu}^{(k+1)}$, $\boldsymbol{\xi}^{(k+1)}$, and $\boldsymbol{\Sigma}^{(k+1)}$.

Based on Lemma 3, we have $p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{q} = \nabla_{\boldsymbol{\mu}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ and $\boldsymbol{q}^\star = \mathbf{0}$. Then the distance (44) becomes

$$\begin{aligned}
\|\boldsymbol{\mu} - \boldsymbol{\mu}^\star\|_2 = \Bigg\| & \underline{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}{p(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}} \nabla_{\underline{\boldsymbol{\mu}}} \log p(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}) \\
& - \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\boldsymbol{\theta}^\star)}{p(\boldsymbol{y}_i;\boldsymbol{\theta}^\star)}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}} \nabla_{\boldsymbol{\mu}^\star} \log p(\boldsymbol{y}_t;\boldsymbol{\theta}^\star) \Bigg\|_2 \\
= & \left\| \underline{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star + \int_0^1 \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}} \nabla^2_{\tilde{\boldsymbol{\mu}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\mu)(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star) \mathrm{d}\beta \right\|_2 \\
\leq & \sup_{\beta\in(0,1]} \left\| \boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}} \nabla^2_{\tilde{\boldsymbol{\mu}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\mu) \right\|_2 \left\| \underline{\boldsymbol{\mu}} - \boldsymbol{\mu}^\star \right\|_2,
\end{aligned}$$

where the second equation is given by the mean value theorem of integrals and $\tilde{\boldsymbol{\theta}}_\mu = \{\tilde{\boldsymbol{\mu}}, \underline{\boldsymbol{\Sigma}}, \underline{\boldsymbol{\xi}}\}$ with $\tilde{\boldsymbol{\mu}} = \beta\underline{\boldsymbol{\mu}} + (1-\beta)\boldsymbol{\mu}^\star$ and $\beta \in (0, 1]$.

To bound the term $\sup_{\beta\in(0,1]} \left\| \boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}} \nabla^2_{\tilde{\boldsymbol{\mu}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\mu) \right\|_2$, we need some results for the second order derivative of $\log p(\boldsymbol{y}_t;\boldsymbol{\theta})$ at first.

Lemma 4. *The second order derivative of* $\log p(\boldsymbol{y};\boldsymbol{\theta})$ *with respect to* $\boldsymbol{\mu}$*, i.e.,* $\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y};\boldsymbol{\theta})$*, satisfy*

1. $\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ *is a negative definite matrix;*
2. $\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\boldsymbol{\theta})}{p(\boldsymbol{y}_i;\boldsymbol{\theta})}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma} \nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}_t;\boldsymbol{\theta})$ *is a positive definite matrix,*

for all $\boldsymbol{\theta}$ *in the feasible set.*

Proof. The second order derivative of $\log p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to $\boldsymbol{\mu}$ is given as

$$\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\mu}} \left(p^{-1}(\boldsymbol{y};\boldsymbol{\theta}) \nabla_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta})\right) = \frac{\nabla^2_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta}) \cdot p(\boldsymbol{y};\boldsymbol{\theta}) - \nabla_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta}) \nabla^\top_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta})}{p^2(\boldsymbol{y};\boldsymbol{\theta})}. \tag{45}$$

To analyze the above expression, we should compute the first and second order derivatives of $p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to $\boldsymbol{\mu}$ at first. The first order one is given as follows:

$$\nabla_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\mu}} \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x};\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \nabla_{\boldsymbol{\mu}} p(\boldsymbol{x};\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}. \tag{46}$$

Since $p(\boldsymbol{x};\boldsymbol{\theta}) = \int_0^{+\infty} p(z) p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \mathrm{d}z$, (46) becomes

$$\begin{aligned}
\nabla_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta}) &= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} p(z) \nabla_{\boldsymbol{\mu}} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} \frac{p(z)}{z} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi}) \mathrm{d}z \mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Then the second order one can be further computed as

$$\begin{aligned}
\nabla^2_{\boldsymbol{\mu}} p(\boldsymbol{y};\boldsymbol{\theta}) &= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} \frac{p(z)}{z} \nabla_{\boldsymbol{\mu}} \left(p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})\right) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} \boldsymbol{\Sigma}^{-1} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) p(z) \left(-z^{-1}\boldsymbol{I} + z^{-2}(\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})^\top \boldsymbol{\Sigma}^{-1}\right) \mathrm{d}z \mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{47}$$

Lemma 5. *For a function $f(\boldsymbol{x}, z)$, if $\int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x} \mid z; \boldsymbol{\theta}) f(\boldsymbol{x}, z) \mathrm{d}\boldsymbol{x}$ is integrable, we have*

$$\int_{\mathbb{R}^d} \int_0^{+\infty} p(z) p(\boldsymbol{x}, \boldsymbol{y} \mid z; \boldsymbol{\theta}) f(\boldsymbol{x}, z) \mathrm{d}z \mathrm{d}\boldsymbol{x} = p(\boldsymbol{y}; \boldsymbol{\theta}) \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[f(\boldsymbol{x}, z) \right].$$

Proof. Given that

$$\int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x} \mid z; \boldsymbol{\theta}) f(\boldsymbol{x}, z) \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^d} p(\boldsymbol{x}, \boldsymbol{y} \mid z; \boldsymbol{\theta}) f(\boldsymbol{x}, z) \mathrm{d}\boldsymbol{x},$$

we can obtain that

$$\begin{aligned} \int_{\mathbb{R}^d} \int_0^{+\infty} p(z) p(\boldsymbol{x}, \boldsymbol{y} \mid z; \boldsymbol{\theta}) f(\boldsymbol{x}, z) \mathrm{d}z \mathrm{d}\boldsymbol{x} &= p(\boldsymbol{y}; \boldsymbol{\theta}) \int_{\mathbb{R}^d} \int_0^{+\infty} p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}) f(\boldsymbol{x}, z) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\ &= p(\boldsymbol{y}; \boldsymbol{\theta}) \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[f(\boldsymbol{x}, z) \right]. \end{aligned}$$

□

Based on Lemma 5, the first and second order derivatives of $p(\boldsymbol{y}; \boldsymbol{\theta})$ becomes

$$\nabla_{\boldsymbol{\mu}} p(\boldsymbol{y}; \boldsymbol{\theta}) = p(\boldsymbol{y}; \boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1} \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} (\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi}) \right],$$

and

$$\nabla^2_{\boldsymbol{\mu}} p(\boldsymbol{y}; \boldsymbol{\theta}) = p(\boldsymbol{y}; \boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1} \left(-\mathsf{E}_{z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \right] \boldsymbol{I} + \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-2} (\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})^\top \right] \boldsymbol{\Sigma}^{-1} \right).$$

Then the second order derivative of $\log p(\boldsymbol{y}; \boldsymbol{\theta})$ in (45) becomes

$$\begin{aligned} \nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}; \boldsymbol{\theta}) &= -\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1} \mathsf{E}_{\boldsymbol{x} \mid \boldsymbol{y}; \boldsymbol{\theta}} [(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top] \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathsf{E}_{\boldsymbol{x} \mid \boldsymbol{y}; \boldsymbol{\theta}} [\boldsymbol{x} - \boldsymbol{\mu}] \mathsf{E}_{\boldsymbol{x} \mid \boldsymbol{y}; \boldsymbol{\theta}} [\boldsymbol{x} - \boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{-1} \\ &= \boldsymbol{\Sigma}^{-1} \left(-\mathsf{E}_{z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \right] \boldsymbol{\Sigma} + \mathsf{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \boldsymbol{x} \right] \right) \boldsymbol{\Sigma}^{-1}. \end{aligned} \tag{48}$$

To prove the negative definite of $\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}; \boldsymbol{\theta})$, we introduce the Brascamp–Lieb inequality.

Lemma 6 (Brascamp & Lieb (1976))**.** *Consider a probability density function $p(\boldsymbol{x})$ which is log-concave to $\boldsymbol{x}$. The Brascamp–Lieb inequality is given by*

$$\mathsf{Cov}_{\tilde{\boldsymbol{x}}}(f(\boldsymbol{x})) \preceq \mathsf{E}_{\tilde{\boldsymbol{x}}} [\nabla_{\boldsymbol{x}}^\top f(\boldsymbol{x}) [\nabla^2_{\boldsymbol{x}} (-\log p(\tilde{\boldsymbol{x}}))]^{-1} \nabla_{\boldsymbol{x}} f(\boldsymbol{x})],$$

where the equality is obtained when $\log p(\tilde{\boldsymbol{x}})$ is linear with respect to $\boldsymbol{x}$.

Let $\tilde{\boldsymbol{x}} = \boldsymbol{x}, z \mid \boldsymbol{y}$ and $f(\boldsymbol{x}) = z^{-1}\boldsymbol{x}$. Since $\nabla^2_{\boldsymbol{x}} \log p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}) = -\frac{\boldsymbol{\Sigma}^{-1}}{z}$, $p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta})$ is a log-concave function with respect to $\boldsymbol{x}$. Based on the Brascamp–Lieb inequality in Lemma 6, since $\log p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta})$ is non-linear to $\boldsymbol{x}$, we have

$$\mathsf{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \boldsymbol{x} \right] \prec \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} z \boldsymbol{\Sigma} z^{-1} \right] = \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \right] \boldsymbol{\Sigma},$$

and hence $\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}; \boldsymbol{\theta})$ is negative definite. Substituting $\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}; \boldsymbol{\theta})$ in (48) into $\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i; \boldsymbol{\theta})}{p(\boldsymbol{y}_i; \boldsymbol{\theta})} \right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma} \nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}_t; \boldsymbol{\theta})$, we have

$$\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i; \boldsymbol{\theta})}{p(\boldsymbol{y}_i; \boldsymbol{\theta})} \right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma} \nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}_t; \boldsymbol{\theta}) = \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1} \mathsf{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}_t; \boldsymbol{\theta}} \left[z^{-1} \boldsymbol{x} \right]}{\sum_{t=1}^n \mathsf{E}_{z \mid \boldsymbol{y}_t; \boldsymbol{\theta}} [z^{-1}]}.$$

Since both $\boldsymbol{\Sigma}$ and $\mathsf{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \boldsymbol{x}_t \right]$ are positive definite matrices, $\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i; \boldsymbol{\theta})}{p(\boldsymbol{y}_i; \boldsymbol{\theta})} \right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma} \nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y}_t; \boldsymbol{\theta})$ is positive definite. □

Based on Lemma 4, we have that all the eigenvalues of matrix $\boldsymbol{I} + \boldsymbol{\Sigma}\nabla^2_{\boldsymbol{\mu}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$ are belongs to $(0,1)$, hence we have

$$\begin{aligned}
&\sup_{\beta\in(0,1]} \left\| \boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\mu)}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}}\nabla^2_{\tilde{\boldsymbol{\mu}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\mu) \right\|_2 \\
= &\sup_{\beta\in(0,1]} \left\| \frac{\sum_{t=1}^n \underline{\boldsymbol{\Sigma}}^{-1}\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\mu}\left[z^{-1}\boldsymbol{x}\right]}{\sum_{t=1}^n \mathsf{E}_{z|\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\mu}[z^{-1}]} \right\|_2 \\
\leq &\max_{\boldsymbol{\theta}} \left\| \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1}\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y}_t;\boldsymbol{\theta}}\left[z^{-1}\boldsymbol{x}\right]}{\sum_{t=1}^n \mathsf{E}_{z|\boldsymbol{y}_t;\boldsymbol{\theta}}[z^{-1}]} \right\|_2 \\
\triangleq &\, c_\mu \in (0,1).
\end{aligned}$$

□

B.2 Convergence Rate of Skewness Parameter

Given the update rule of $\boldsymbol{\xi}$ in (12), we have

$$\boldsymbol{\xi} = \frac{\sum_{t=1}^n (\boldsymbol{w}_t - \boldsymbol{\mu})}{\sum_{t=1}^n \zeta_t} = \frac{\sum_{t=1}^n (\mathsf{E}_{\boldsymbol{x}|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[\boldsymbol{x}] - \underline{\boldsymbol{\mu}})}{\sum_{t=1}^n \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z]}. \tag{49}$$

Based on (38), we have

$$\mathsf{E}_{\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}\right] - \boldsymbol{\mu} = \mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}[z]\boldsymbol{\xi} + p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{q}_1.$$

Hence, the update rule of $\boldsymbol{\xi}$ in (49) becomes

$$\boldsymbol{\xi} = \underline{\boldsymbol{\xi}} + \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}{p(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}\right)^{-1} \sum_{t=1}^n p^{-1}(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}\boldsymbol{q}_{1,t} \tag{50}$$

Hence, the distance between $\boldsymbol{\xi}$ at the current iteration and $\boldsymbol{\xi}^\star$ is given by

$$\|\boldsymbol{\xi} - \boldsymbol{\xi}^\star\|_2 = \left\| \underline{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star + \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}{p(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}\right)^{-1} \sum_{t=1}^n p^{-1}(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}\boldsymbol{q}_{1,t} \right\|_2. \tag{51}$$

To further analyze the term on the right side in (51), we first give a result of $\boldsymbol{q}_1$.

Lemma 7. *When $\boldsymbol{x}$ follows a normal mean-variance mixture and $\boldsymbol{q}_1$ is defined in (21), we have*

$$\boldsymbol{q}_1 = \nabla_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}).$$

Proof. Consider the partial derivative of $p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to the i-th element of $\boldsymbol{\xi}$. We have that

$$\nabla_{\xi_i} p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\xi_i} \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x} = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \nabla_{\xi_i} p(\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}.$$

Since the partial derivatives of $p(\boldsymbol{x};\boldsymbol{\theta})$

$$\nabla_{\xi_i} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) = -z\nabla_{x_i} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}),$$

we can obtain that

$$\begin{aligned}
\nabla_{\xi_i} p(\boldsymbol{y};\boldsymbol{\theta}) &= -\int_0^{+\infty} \int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i})} \int_{l_i}^{u_i} z\nabla_{x_i} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) p(z)\mathrm{d}x_i\mathrm{d}\boldsymbol{x}_{\setminus i}\mathrm{d}z \\
&= \int_0^{+\infty} \int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i})} \frac{p(z)}{z} z\left(p(\boldsymbol{x}_{\setminus i}, x_i = l_i \mid z;\boldsymbol{\theta}) - p(\boldsymbol{x}_{\setminus i}, x_i = u_i \mid z;\boldsymbol{\theta})\right)\mathrm{d}\boldsymbol{x}_{\setminus i}\mathrm{d}z \\
&= p(x_i = l_i, \boldsymbol{y}_{\setminus i} \mid z,\boldsymbol{\theta}) - p(x_i = u_i, \boldsymbol{y}_{\setminus i} \mid z,\boldsymbol{\theta}).
\end{aligned}$$

which is equivalent to the definition of the i-th element of $\boldsymbol{q}_1$ from (33).

Based on Lemma 7, we have $p^{-1}(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{q}_1 = \nabla_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ and $\boldsymbol{q}_1^\star = \mathbf{0}$. Then the distance (44) becomes

$$\begin{aligned}
\|\boldsymbol{\xi} - \boldsymbol{\xi}^\star\|_2 = \Bigg\| & \underline{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star + \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}{p(\boldsymbol{y}_i;\underline{\boldsymbol{\theta}})}\right)^{-1} \sum_{t=1}^n \underline{\boldsymbol{\Sigma}} \nabla_{\underline{\boldsymbol{\xi}}} \log p(\boldsymbol{y};\underline{\boldsymbol{\theta}}) \\
& - \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\boldsymbol{\theta}^\star)}{p(\boldsymbol{y}_i;\boldsymbol{\theta}^\star)}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma}^\star \nabla_{\boldsymbol{\xi}^\star} \log p(\boldsymbol{y};\boldsymbol{\theta}^\star) \Bigg\|_2 \\
= & \left\| \underline{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star + \int_0^1 \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}\right)^{-1} \sum_{t=1}^n \tilde{\boldsymbol{\Sigma}} \nabla^2_{\tilde{\boldsymbol{\xi}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\xi)(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star) \mathrm{d}\beta \right\|_2 \\
\leq & \sup_{\beta\in(0,1]} \left\| \boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}\right)^{-1} \sum_{t=1}^n \tilde{\boldsymbol{\Sigma}} \nabla^2_{\tilde{\boldsymbol{\xi}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\xi) \right\|_2 \left\| \underline{\boldsymbol{\xi}} - \boldsymbol{\xi}^\star \right\|_2,
\end{aligned}$$

where the second equation is given by the mean value theorem of integrals and $\tilde{\boldsymbol{\theta}}_\xi = \left\{\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{\xi}}\right\}$ with $\tilde{\boldsymbol{\xi}} = \beta\underline{\boldsymbol{\xi}} + (1-\beta)\boldsymbol{\xi}^\star$, and $\beta \in (0, 1]$.

To bound the term $\sup_{\beta\in(0,1]} \left\| \boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}\right)^{-1} \sum_{t=1}^n \tilde{\boldsymbol{\Sigma}} \nabla^2_{\tilde{\boldsymbol{\xi}}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\xi) \right\|_2$, we need some results for the second order derivative of $\log p(\boldsymbol{y}_t;\boldsymbol{\theta})$ at first.

Lemma 8. *The second order derivative of* $\log p(\boldsymbol{y};\boldsymbol{\theta})$ *with respect to* $\boldsymbol{\xi}$*, i.e.,* $\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta})$*, satisfy*

1. $\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ *is a negative definite matrix;*
2. $\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\boldsymbol{\theta})}{p(\boldsymbol{y}_i;\boldsymbol{\theta})}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma} \nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y}_t;\boldsymbol{\theta})$ *is a positive definite matrix.*

Proof. The second order derivative of $\log p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to $\boldsymbol{\xi}$ is given as

$$\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\xi}} \left(p^{-1}(\boldsymbol{y};\boldsymbol{\theta}) \nabla_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta})\right) = \frac{\nabla^2_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) \cdot p(\boldsymbol{y};\boldsymbol{\theta}) - \nabla_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) \nabla^\top_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta})}{p^2(\boldsymbol{y};\boldsymbol{\theta})}. \tag{52}$$

To analyze the above expression, we should compute the first and second order derivatives of $p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to $\boldsymbol{\xi}$ at first. The first order one is given as follows:

$$\nabla_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\xi}} \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x};\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \nabla_{\boldsymbol{\xi}} p(\boldsymbol{x};\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}. \tag{53}$$

Since $p(\boldsymbol{x};\boldsymbol{\theta}) = \int_0^{+\infty} p(z) p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \mathrm{d}z$, (53) becomes

$$\begin{aligned}
\nabla_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) &= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} p(z) \nabla_{\boldsymbol{\xi}} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} p(z) p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi}) \mathrm{d}z \mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Then the second order derivative can be further computed as

$$\begin{aligned}
\nabla^2_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) &= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} p(z) \nabla_{\boldsymbol{\xi}} \left(p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})\right) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} \boldsymbol{\Sigma}^{-1} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) p(z) \left(-z\boldsymbol{I} + (\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{\mu} - z\boldsymbol{\xi})^\top \boldsymbol{\Sigma}^{-1}\right) \mathrm{d}z \mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{54}$$

Based on Lemma 5, the derivatives of $p(\boldsymbol{y};\boldsymbol{\theta})$ becomes

$$\nabla_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) = p(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}-\boldsymbol{\mu}-z\boldsymbol{\xi}\right],$$

and

$$\nabla^2_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta}) = p(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}\left(-\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z\right]\boldsymbol{I} + \mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[(\boldsymbol{x}-\boldsymbol{\mu}-z\boldsymbol{\xi})(\boldsymbol{x}-\boldsymbol{\mu}-z\boldsymbol{\xi})^\top\right]\boldsymbol{\Sigma}^{-1}\right).$$

Then the second order derivative of $\log p(\boldsymbol{y};\boldsymbol{\theta})$ in (52) becomes

$$\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1}\left(-\mathsf{E}_{z|\boldsymbol{y};\boldsymbol{\theta}}\left[z\right]\boldsymbol{\Sigma} + \mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}-z\boldsymbol{\xi}\right]\right)\boldsymbol{\Sigma}^{-1}. \tag{55}$$

Setting $\tilde{\boldsymbol{x}} = \boldsymbol{x}, z \mid \boldsymbol{y}$ and $f(\boldsymbol{x}) = \boldsymbol{x} - z\boldsymbol{\xi}$, since $\nabla^2_{\boldsymbol{x}} \log p(\boldsymbol{x}, z \mid \boldsymbol{y};\boldsymbol{\theta}) = -\frac{\boldsymbol{\Sigma}^{-1}}{z}$, $p(\boldsymbol{x}, z \mid \boldsymbol{y};\boldsymbol{\theta})$ is a log-concave function with respect to $\boldsymbol{x}$. Based on the Brascamp–Lieb inequality in Lemma 6, we have

$$\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}-z\boldsymbol{\xi}\right] \prec \mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[z\boldsymbol{\Sigma}\right] = \mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[z\right]\boldsymbol{\Sigma},$$

and hence $\nabla^2_{\boldsymbol{\xi}} p(\boldsymbol{y};\boldsymbol{\theta})$ is negative definite.

Substituting $\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ in (55) into $\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\boldsymbol{\theta})}{p(\boldsymbol{y}_i;\boldsymbol{\theta})}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma}\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y}_t;\boldsymbol{\theta})$, we have

$$\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\boldsymbol{\theta})}{p(\boldsymbol{y}_i;\boldsymbol{\theta})}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma}\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y}_t;\boldsymbol{\theta}) = \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1}\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y}_t;\boldsymbol{\theta}}\left[\boldsymbol{x}-z\boldsymbol{\xi}\right]}{\sum_{t=1}^n \mathsf{E}_{z|\boldsymbol{y}_t;\boldsymbol{\theta}}[z]}.$$

Since both $\boldsymbol{\Sigma}$ and $\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\boldsymbol{x}_t - z\boldsymbol{\xi}\right]$ are positive definite matrices, $\boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_{-1}(\boldsymbol{y}_i;\boldsymbol{\theta})}{p(\boldsymbol{y}_i;\boldsymbol{\theta})}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma}\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y}_t;\boldsymbol{\theta})$ is positive definite. □

Based on Lemma 8, we have that all the eigenvalues of matrix $\boldsymbol{I} + \boldsymbol{\Sigma}\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y};\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$ are belong to $(0,1)$, hence we have

$$\begin{aligned}
&\sup_{\beta\in(0,1]} \left\| \boldsymbol{I} + \left(\sum_{i=1}^n \frac{p_1(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}{p(\boldsymbol{y}_i;\tilde{\boldsymbol{\theta}}_\xi)}\right)^{-1} \sum_{t=1}^n \boldsymbol{\Sigma}\nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\xi) \right\|_2 \\
=&\sup_{\beta\in(0,1]} \left\| \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1}\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\xi}\left[\boldsymbol{x}-z\boldsymbol{\xi}\right]}{\sum_{t=1}^n \mathsf{E}_{z|\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_\xi}[z]} \right\|_2 \\
\le&\max_{\boldsymbol{\theta}} \left\| \frac{\sum_{t=1}^n \boldsymbol{\Sigma}^{-1}\mathsf{Cov}_{\boldsymbol{x},z|\boldsymbol{y}_t;\boldsymbol{\theta}}\left[\boldsymbol{x}-z\boldsymbol{\xi}\right]}{\sum_{t=1}^n \mathsf{E}_{z|\boldsymbol{y}_t;\boldsymbol{\theta}}[z]} \right\|_2 \\
\triangleq& c_\xi \in (0,1).
\end{aligned}$$

□

B.3 Convergence Rate of Scatter Matrix

Given the update rule of $\boldsymbol{\Sigma}$ in (12), we have

$$\begin{aligned}
\boldsymbol{\Sigma} &= \frac{1}{n}\sum_{t=1}^n \left(\left(\boldsymbol{U}_t - 2\boldsymbol{v}_t\underline{\boldsymbol{\mu}}^\top + \iota_t\underline{\boldsymbol{\mu}}\underline{\boldsymbol{\mu}}^\top\right) - 2(\boldsymbol{w}_t - \underline{\boldsymbol{\mu}})\underline{\boldsymbol{\xi}}^\top + \zeta_t\underline{\boldsymbol{\xi}}\underline{\boldsymbol{\xi}}^\top\right) \\
&= \frac{1}{n}\sum_{t=1}^n \left(\mathsf{E}_{\boldsymbol{x}_t,z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z^{-1}\boldsymbol{x}\boldsymbol{x}^\top] - 2\mathsf{E}_{\boldsymbol{x}_t,z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z^{-1}\boldsymbol{x}_t]\underline{\boldsymbol{\mu}}^\top + \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z^{-1}]\underline{\boldsymbol{\mu}}\underline{\boldsymbol{\mu}}^\top\right. \\
&\quad \left. -2(\mathsf{E}_{\boldsymbol{x}|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[\boldsymbol{x}] - \underline{\boldsymbol{\mu}})\underline{\boldsymbol{\xi}}^\top + \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z]\underline{\boldsymbol{\xi}}\underline{\boldsymbol{\xi}}^\top\right).
\end{aligned} \tag{56}$$

In the following, we derive all five terms in the summation in (56).

Term 1: Based on (41), we have

$$\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\underline{\boldsymbol{\theta}}}\left[z^{-1}\boldsymbol{x}\boldsymbol{x}^\top\right] = \mathsf{E}_{z|\boldsymbol{y};\underline{\boldsymbol{\theta}}}\left[z^{-1}\right]\underline{\boldsymbol{\mu}}\underline{\boldsymbol{\mu}}^\top + \mathsf{E}_{z|\boldsymbol{y};\underline{\boldsymbol{\theta}}}\left[z\right]\underline{\boldsymbol{\xi}}\underline{\boldsymbol{\xi}}^\top + 2\underline{\boldsymbol{\mu}}\underline{\boldsymbol{\xi}}^\top + \underline{\boldsymbol{\Sigma}} + p^{-1}(\boldsymbol{y};\underline{\boldsymbol{\theta}})\left(2\left(\underline{\boldsymbol{\Sigma}}\boldsymbol{q}\right)\underline{\boldsymbol{\mu}}^\top + 2\mathsf{E}_z\left[z\right]\left(\underline{\boldsymbol{\Sigma}}\boldsymbol{q}_1\right)\underline{\boldsymbol{\xi}}^\top + \underline{\boldsymbol{\Sigma}}\left(\boldsymbol{H}_1 + \boldsymbol{D}_1\right)\underline{\boldsymbol{\Sigma}}\right),$$

where $\boldsymbol{q}$, $\boldsymbol{q}_1$, $\boldsymbol{H}_1$ and $\boldsymbol{D}_1$ are defined in (21), (32), (34) and (37), respectively.

Term 2: Based on (40), we can obtain

$$-2\mathsf{E}_{\boldsymbol{x}_t,z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[z_t^{-1}\boldsymbol{x}_t]\underline{\boldsymbol{\mu}}^\top = -2\left(\mathsf{E}_{z|\boldsymbol{y};\underline{\boldsymbol{\theta}}}\left[z^{-1}\right]\underline{\boldsymbol{\mu}}\underline{\boldsymbol{\mu}}^\top + \boldsymbol{\xi} + p^{-1}(\boldsymbol{y};\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}\boldsymbol{q}\right)\underline{\boldsymbol{\mu}}^\top.$$

Term 4: Based on (38), we have

$$-2(\mathsf{E}_{\boldsymbol{x}_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}[\boldsymbol{x}_t] - \underline{\boldsymbol{\mu}})\underline{\boldsymbol{\xi}}^\top = -2\mathsf{E}_{z|\boldsymbol{y};\underline{\boldsymbol{\theta}}}\left[z\right]\underline{\boldsymbol{\xi}}\underline{\boldsymbol{\xi}}^\top - 2p^{-1}(\boldsymbol{y};\underline{\boldsymbol{\theta}})\left(\underline{\boldsymbol{\Sigma}}\boldsymbol{q}_1\right)\underline{\boldsymbol{\xi}}^\top$$

Upon cancellation of the opposing terms, the update rule in (56) reduces to:

$$\boldsymbol{\Sigma} = \underline{\boldsymbol{\Sigma}} + \frac{1}{n}\sum_{t=1}^{n} p^{-1}(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}(\boldsymbol{H}_{1,t} + \boldsymbol{D}_{1,t})\underline{\boldsymbol{\Sigma}}, \tag{57}$$

Hence, the distance between $\boldsymbol{\Sigma}$ at the current iteration and $\boldsymbol{\Sigma}^\star$ is given by

$$\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\|_{\mathsf{F}} = \left\|\underline{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star + \frac{1}{n}\sum_{t=1}^{n} p^{-1}(\boldsymbol{y}_t;\underline{\boldsymbol{\theta}})\underline{\boldsymbol{\Sigma}}(\boldsymbol{H}_{1,t} + \boldsymbol{D}_{1,t})\underline{\boldsymbol{\Sigma}}\right\|_{\mathsf{F}}. \tag{58}$$

To further analyze the term on the right side, we first give a result of $\boldsymbol{H}_{1,t} + \boldsymbol{D}_{1,t}$.

Lemma 9. *When $\boldsymbol{x}$ follows a normal mean variance mixture and $\boldsymbol{H}_1$ and $\boldsymbol{D}_1$ are defined in (34) and (37), respectively, we have*

$$\frac{1}{2}(\boldsymbol{H}_1 + \boldsymbol{D}_1) = \nabla_{\boldsymbol{\Sigma}} p(\boldsymbol{y};\boldsymbol{\theta}).$$

Proof. Considering the partial derivative of $p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to the (i,j)-th element of $\boldsymbol{\Sigma}$, we have that

$$\nabla_{\boldsymbol{\Sigma}_{ij}} p(\boldsymbol{y};\boldsymbol{\theta}) = \nabla_{\boldsymbol{\Sigma}_{ij}} \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} p(\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x} = \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \nabla_{\boldsymbol{\Sigma}_{ij}} p(\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}.$$

The partial derivatives of $p(\boldsymbol{x};\boldsymbol{\theta})$ satisfy the following equation:

$$\nabla_{\boldsymbol{\Sigma}_{ij}} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}) = \frac{z}{2}\nabla^2_{x_i x_j} p(\boldsymbol{x} \mid z;\boldsymbol{\theta}), \tag{59}$$

Hence, for $i \neq j$, we can obtain that

$$\begin{aligned}
\nabla_{\boldsymbol{\Sigma}_{ij}} p(\boldsymbol{y};\boldsymbol{\theta}) &= \int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i,j})} \int_{l_i}^{u_i}\int_{l_j}^{u_j}\int_0^{+\infty} \frac{z}{2}\nabla^2_{x_i x_j} p(\boldsymbol{x} \mid z;\boldsymbol{\theta})p(z)\mathrm{d}z\mathrm{d}x_i\mathrm{d}x_j\mathrm{d}\boldsymbol{x}_{\setminus i,j} \\
&= \int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i,j})} \int_0^{+\infty} \frac{zp(z)}{2}\left(p(\boldsymbol{x}_{\setminus i,j}, x_i = l_i, x_j = l_j \mid z;\boldsymbol{\theta}) - p(\boldsymbol{x}_{\setminus i,j}, x_i = l_i, x_j = u_j \mid z;\boldsymbol{\theta})\right. \\
&\quad \left. - p(\boldsymbol{x}_{\setminus i,j}, x_i = u_i, x_j = l_j \mid z;\boldsymbol{\theta}) + p(\boldsymbol{x}_{\setminus i,j}, x_i = u_i, x_j = u_j \mid z;\boldsymbol{\theta})\right)\mathrm{d}z\mathrm{d}\boldsymbol{x}_{\setminus i,j} \\
&= \frac{1}{2}\mathsf{E}_z(z)\left(p_1(\boldsymbol{y}_{\setminus i,j}, x_i = l_i, x_j = l_j;\boldsymbol{\theta}) - p_1(\boldsymbol{y}_{\setminus i,j}, x_i = l_i, x_j = u_j;\boldsymbol{\theta})\right. \\
&\quad \left. - p_1(\boldsymbol{y}_{\setminus i,j}, x_i = u_i, x_j = l_j;\boldsymbol{\theta}) + p_1(\boldsymbol{y}_{\setminus i,j}, x_i = u_i, x_j = u_j;\boldsymbol{\theta})\right) \\
&= \frac{1}{2}\boldsymbol{H}_{1,ij}.
\end{aligned}$$

For the case $i = j$, we first introduce a result.

Lemma 10. *When $\boldsymbol{x}$ follows a normal mean-variance mixture, we have the following equation:*

$$\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik}\nabla^2_{x_k x_i} p_1(\boldsymbol{x} \mid z;\boldsymbol{\theta}) = \nabla_{x_i}\left(-\frac{p_1(\boldsymbol{x} \mid z;\boldsymbol{\theta})}{z}(x_i - \mu_i - z\xi_i)\right). \tag{60}$$

Proof. We begin with

$$\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \nabla^2_{x_k x_i} p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta}) = \nabla_{x_i} \left(\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \nabla_{x_k} p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta}) \right)$$

For the term $\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \nabla_{x_k} p(\boldsymbol{x} \mid z; \boldsymbol{\theta})$, we have

$$\begin{aligned} \sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \nabla_{x_k} p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta}) &= \sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \left(-\frac{p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta})}{z} \sum_{l=1}^{d} \boldsymbol{\Sigma}^{-1}_{kl} (x_l - \mu_l - z\xi_l) \right) \\ &= -\frac{p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta})}{z} \sum_{l=1}^{d} \left(\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \boldsymbol{\Sigma}^{-1}_{kl} (x_l - \mu_l - z\xi_l) \right). \end{aligned}$$

Since $\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \boldsymbol{\Sigma}^{-1}_{kl}$ is equivalent to the (i, l)-entry of the matrix $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}$, we can obtain that

$$\sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \boldsymbol{\Sigma}^{-1}_{kl} = \begin{cases} 0, & i \neq l, \\ 1, & i = l. \end{cases}$$

Hence, we have

$$-\frac{p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta})}{z} \sum_{k=1}^{d} \left(\sum_{l=1}^{d} \boldsymbol{\Sigma}_{ik} \boldsymbol{\Sigma}^{-1}_{kl} (x_l - \mu_l - z\xi_l) \right) = -\frac{p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta})}{z} (x_i - \mu_i - z\xi_i).$$

Therefore, we prove the equation (60). □

Computing the integral of the left side of (60), we have

$$\begin{aligned} &\int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} \sum_{k=1}^{d} \boldsymbol{\Sigma}_{ik} \nabla^2_{x_k x_i} p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta}) p(z) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\ &= \int_0^{+\infty} \int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \left(\sigma_i \nabla^2_{x_i} p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta}) + \sum_{k \neq i} \boldsymbol{\Sigma}_{ik} \nabla^2_{x_k x_i} p_1(\boldsymbol{x} \mid z; \boldsymbol{\theta}) \right) p(z) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\ &= 2\mathsf{E}_z[z^{-1}] \sigma_i \nabla_{\sigma_i} p(\boldsymbol{y}; \boldsymbol{\theta}) + [\boldsymbol{\Sigma}\boldsymbol{H}_1]_{ii}. \end{aligned} \tag{61}$$

Then the integral of the right side of (60) is given by

$$\begin{aligned} &\int_{\mathcal{Q}^{-1}(\boldsymbol{y})} \int_0^{+\infty} \nabla_{x_i} \left(-\frac{p_1(\boldsymbol{x}; \boldsymbol{\theta})}{z} (x_i - \mu_i - z\xi_i) \right) \mathrm{d}z \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{Q}^{-1}(\boldsymbol{y}_{\setminus i})} \int_{l_i}^{u_i} \int_0^{+\infty} \nabla_{x_i} \left(-\frac{p_1(\boldsymbol{x}; \boldsymbol{\theta})}{z} (x_i - \mu_i - z\xi_i) \right) \mathrm{d}z \mathrm{d}x_i \mathrm{d}\boldsymbol{x}_{\setminus i} \\ &= \mathsf{E}_z(z^{-1}) \left((l_i - \mu_i) p(x_i = l_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) - \xi_i p_1(x_i = l_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) \right. \\ &\quad \left. -(u_i - \mu_i) p(x_i = u_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) + \xi_i p_1(x_i = u_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) \right). \end{aligned} \tag{62}$$

Since (61) and (62) are equivalent, we have

$$\begin{aligned} \nabla_{\sigma_i} p(\boldsymbol{y}; \boldsymbol{\theta}) = \frac{1}{2\sigma_i} &\left((l_i - \mu_i) p(x_i = l_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) - \xi_i p_1(x_i = l_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) \right. \\ &\left. -(u_i - \mu_i) p(x_i = u_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) + \xi_i p_1(x_i = u_i, \boldsymbol{y}_{\setminus i}; \boldsymbol{\theta}) - \mathsf{E}_z[z][\boldsymbol{\Sigma}\boldsymbol{H}_1]_{ii} \right) = \frac{1}{2} \boldsymbol{D}_{1,ii}. \end{aligned}$$

Therefore, $\frac{1}{2}(\boldsymbol{H}_1 + \boldsymbol{D}_1) = \nabla_{\boldsymbol{\Sigma}} p(\boldsymbol{y}; \boldsymbol{\theta})$ is valid. □

Based on Lemma 9, we have that

$$p^{-1}(\boldsymbol{y}; \boldsymbol{\theta}) \boldsymbol{\Sigma} (\boldsymbol{H}_1 + \boldsymbol{D}_1) \boldsymbol{\Sigma} = -2 \nabla_{\boldsymbol{\Sigma}^{-1}} \log p(\boldsymbol{y}; \boldsymbol{\theta}).$$

Hence, the distance (58) becomes

$$\begin{aligned}\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}^{\star}\|_{\mathsf{F}} &= \left\|\underline{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}^{\star}-\frac{2}{n}\sum_{t=1}^{n}\nabla_{\underline{\boldsymbol{\Sigma}}^{-1}}\log p(\boldsymbol{y};\underline{\boldsymbol{\theta}})+\frac{2}{n}\sum_{t=1}^{n}\nabla_{\boldsymbol{\Sigma}^{\star-1}}\log p(\boldsymbol{y};\boldsymbol{\theta}^{\star})\right\|_{\mathsf{F}} \\ &= \left\|\underline{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}^{\star}-\frac{2}{n}\sum_{t=1}^{n}\int_{0}^{1}\nabla^{2}_{\tilde{\boldsymbol{\Sigma}}^{-1},\tilde{\boldsymbol{\Sigma}}}\log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_{\Sigma})(\tilde{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}^{\star})\mathrm{d}\beta\right\|_{\mathsf{F}} \\ &\leq \sup_{\beta\in(0,1]}\left\|\boldsymbol{I}_{d\times d}-\frac{2}{n}\sum_{t=1}^{n}\frac{\partial\mathrm{vec}\left(\nabla_{\tilde{\boldsymbol{\Sigma}}^{-1}}\log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_{\Sigma})\right)}{\partial\mathrm{vec}\left(\tilde{\boldsymbol{\Sigma}}\right)}\right\|_{2}\|\underline{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}^{\star}\|_{\mathsf{F}}.\end{aligned} \tag{63}$$

where the second equation is given by the mean value theorem of integrals and $\tilde{\boldsymbol{\theta}}_{\Sigma}=\left\{\underline{\boldsymbol{\mu}},\tilde{\boldsymbol{\Sigma}},\underline{\boldsymbol{\xi}}\right\}$ with $\tilde{\boldsymbol{\Sigma}}=\beta\underline{\boldsymbol{\Sigma}}+(1-\beta)\boldsymbol{\Sigma}^{\star}$ and $\beta\in(0,1]$.

To bound the term $\sup_{\beta\in(0,1]}\left\|\boldsymbol{I}_{d\times d}-\frac{2}{n}\sum_{t=1}^{n}\frac{\partial\mathrm{vec}\left(\nabla_{\tilde{\boldsymbol{\Sigma}}^{-1}}\log p(\boldsymbol{y}_t;\tilde{\boldsymbol{\theta}}_{\Sigma})\right)}{\partial\mathrm{vec}\left(\tilde{\boldsymbol{\Sigma}}\right)}\right\|_{2}$, we need some results for the term $\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}\log p(\boldsymbol{y}_t;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma})}$ at first.

Lemma 11. *The term* $\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}\log p(\boldsymbol{y};\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma})}$ *satisfy*

1. *it is a positive definite matrix;*
2. $\boldsymbol{I}_{d\times d}-\frac{2}{n}\sum_{t=1}^{n}\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}\log p(\boldsymbol{y}_t;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma})}$ *is also a positive definite matrix.*

Proof. Since $\log p(\boldsymbol{x}\mid z;\boldsymbol{\theta})$ is linear with respect to $\boldsymbol{\Sigma}^{-1}$, the second order derivative of $\log p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to $\boldsymbol{\Sigma}^{-1}$ is easier to be obtained. We first compute $\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}\log p(\boldsymbol{y}_t;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}^{-1})}$ and transform it later.

The derivative term $\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}\log p(\boldsymbol{y}_t;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}^{-1})}$ can be computed as

$$\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}\log p(\boldsymbol{y};\boldsymbol{\theta})\right)}{\partial\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)}=\frac{\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y}_t;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}^{-1})}\cdot p(\boldsymbol{y};\boldsymbol{\theta})-\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})\otimes\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})}{p^{2}(\boldsymbol{y};\boldsymbol{\theta})}. \tag{64}$$

Since the first and second order derivatives of $p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to $\boldsymbol{\Sigma}^{-1}$ satisfy

$$\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})=\int_{\mathcal{Q}^{-1}(\boldsymbol{y})}\int_{0}^{+\infty}\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})\mathrm{d}z\mathrm{d}\boldsymbol{x},$$

and

$$\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})\right)}{\partial\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)}=\int_{\mathcal{Q}^{-1}(\boldsymbol{y})}\int_{0}^{+\infty}\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)}\mathrm{d}z\mathrm{d}\boldsymbol{x},$$

based on Lemma 5, the expression (64) becomes

$$\begin{aligned}&\frac{\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}^{-1})}\cdot p(\boldsymbol{y};\boldsymbol{\theta})-\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})\otimes\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{y};\boldsymbol{\theta})}{p^{2}(\boldsymbol{y};\boldsymbol{\theta})} \\ =&\frac{\int_{\mathcal{Q}^{-1}(\boldsymbol{y})}\int_{0}^{+\infty}\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}^{-1})}\mathrm{d}z\mathrm{d}\boldsymbol{x}}{p(\boldsymbol{y};\boldsymbol{\theta})} \\ &-\frac{\int_{\mathcal{Q}^{-1}(\boldsymbol{y})}\int_{0}^{+\infty}\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})\mathrm{d}z\mathrm{d}\boldsymbol{x}}{p(\boldsymbol{y};\boldsymbol{\theta})}\otimes\frac{\int_{\mathcal{Q}^{-1}(\boldsymbol{y})}\int_{0}^{+\infty}\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})\mathrm{d}z\mathrm{d}\boldsymbol{x}}{p(\boldsymbol{y};\boldsymbol{\theta})} \\ =&\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\frac{\frac{\partial\mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})\right)}{\partial\mathrm{vec}(\boldsymbol{\Sigma}^{-1})}}{p(\boldsymbol{x},z;\boldsymbol{\theta})}\right]-\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\frac{\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})}{p(\boldsymbol{x},z;\boldsymbol{\theta})}\right]\otimes\mathsf{E}_{\boldsymbol{x},z|\boldsymbol{y};\boldsymbol{\theta}}\left[\frac{\nabla_{\boldsymbol{\Sigma}^{-1}}p(\boldsymbol{x},z;\boldsymbol{\theta})}{p(\boldsymbol{x},z;\boldsymbol{\theta})}\right].\end{aligned} \tag{65}$$

Since we have

$$\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta}) = \frac{p(\boldsymbol{x}, z; \boldsymbol{\theta})}{2z} \left(\boldsymbol{\Sigma} - (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right),$$

and

$$\begin{aligned} \frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)} &= \frac{p(\boldsymbol{x}, z; \boldsymbol{\theta})}{4z^2} \left(\boldsymbol{\Sigma} - (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right) \otimes \left(\boldsymbol{\Sigma} - (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right) \\ &\quad - \frac{p(\boldsymbol{x}, z; \boldsymbol{\theta})}{2z} \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}, \end{aligned}$$

the expression (65) can further be derived as

$$\begin{aligned} &\mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[\frac{\frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)}}{p(\boldsymbol{x}, z; \boldsymbol{\theta})} \right] - \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[\frac{\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})}{p(\boldsymbol{x}, z; \boldsymbol{\theta})} \right] \otimes \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[\frac{\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})}{p(\boldsymbol{x}, z; \boldsymbol{\theta})} \right] \\ &= \frac{1}{4} \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[\frac{1}{z^2} \mathrm{vec}\left((\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right) \mathrm{vec}\left((\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right)^\top - \frac{1}{2z} \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right] \\ &\quad - \frac{1}{4} \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \mathrm{vec}\left((\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right) \right] \mathsf{E}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \mathrm{vec}\left((\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \right)^\top \right] \\ &= \frac{1}{4} \mathrm{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \mathrm{vec}\left(\boldsymbol{x} \boldsymbol{x}^\top \right) \right] - \frac{1}{2} \mathsf{E}_{z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right]. \end{aligned} \tag{66}$$

Setting $\tilde{\boldsymbol{x}} = \boldsymbol{x}, z \mid \boldsymbol{y}$ and $f(\mathrm{vec}\left(\boldsymbol{x}\boldsymbol{x}^\top\right)) = z^{-1} \mathrm{vec}\left(\boldsymbol{x}\boldsymbol{x}^\top\right)$, since $\nabla^2_{\mathrm{vec}(\boldsymbol{x}\boldsymbol{x}^\top)} \log p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}) = -\frac{\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}}{z}$, $p(\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta})$ is a log-concave function with respect to $\mathrm{vec}\left(\boldsymbol{x}\boldsymbol{x}^\top\right)$. Based on the Brascamp–Lieb inequality in Lemma 6, we have

$$\mathrm{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left(z^{-1} \mathrm{vec}\left(\boldsymbol{x} \boldsymbol{x}^\top \right) \right) \prec 2 \mathsf{E}_{z \mid \boldsymbol{y}; \boldsymbol{\theta}} \left[z^{-1} \right] \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}.$$

Since $\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$ is positive definite, $\frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{y}; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)}$ is negative definite.

Based on

$$\frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}\right)} = - \frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right),$$

we can obtain that $\frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{x}, z; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}\right)}$ is positive definite and

$$\begin{aligned} &\boldsymbol{I}_{d \times d} - \frac{2}{n} \sum_{t=1}^{n} \frac{\partial \mathrm{vec}\left(\nabla_{\boldsymbol{\Sigma}^{-1}} p(\boldsymbol{y}_t; \boldsymbol{\theta})\right)}{\partial \mathrm{vec}\left(\boldsymbol{\Sigma}\right)} \\ &= \frac{1}{2n} \sum_{t=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \mathrm{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}_t; \boldsymbol{\theta}} \left(\mathrm{vec}\left(\boldsymbol{x} \boldsymbol{x}^\top \right) \right) \succ \boldsymbol{0}. \end{aligned}$$

□

Based on Lemma 11, we have

$$\begin{aligned} &\sup_{\beta \in (0,1]} \left\| \boldsymbol{I}_{d \times d} - \sum_{t=1}^{n} \frac{\partial \mathrm{vec}\left(\nabla_{\tilde{\boldsymbol{\Sigma}}^{-1}} \log p(\boldsymbol{y}_t; \tilde{\boldsymbol{\theta}}_\Sigma) \right)}{\partial \mathrm{vec}\left(\tilde{\boldsymbol{\Sigma}} \right)} \right\|_2 \\ &= \sup_{\beta \in (0,1]} \left\| \frac{1}{2n} \sum_{t=1}^{n} \left(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \tilde{\boldsymbol{\Sigma}}^{-1} \right) \mathrm{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}_t; \tilde{\boldsymbol{\theta}}_\Sigma} \left[\mathrm{vec}\left(\boldsymbol{x} \boldsymbol{x}^\top \right) \right] \right\|_2 \\ &= \max_{\boldsymbol{\theta}} \left\| \frac{1}{2n} \sum_{t=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \mathrm{Cov}_{\boldsymbol{x}, z \mid \boldsymbol{y}_t; \boldsymbol{\theta}} \left(\mathrm{vec}\left(\boldsymbol{x} \boldsymbol{x}^\top \right) \right) \right\|_2 \\ &\triangleq c_\Sigma \in (0, 1). \end{aligned}$$

C Proof of Theorem 2

Based on Proposition 1, we have proved the linear convergence rates of the three parameters in $\boldsymbol{\theta}$. To connect the global convergence rates and the separable convergence rates, we introduce a result.

Lemma 12 (Meng & Rubin (1994)). *The global convergence rate of an ECM algorithm is the maximum of the componentwise rates of convergence.*

Based on Lemma 12, we obtain that

$$\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{\star}\|_2 \leq c^k \|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^{\star}\|_2,$$

where $c = \max\{c_\mu, c_\xi, c_\Sigma\} \in (0, 1)$.

D Derivations of Surrogate Functions in Quantized Matrix Completion and Compressive Sensing

D.1 Surrogate Function Derivation in Quantized Matrix Completion

The goal of the low-rank matrix completion problem is to recover an unknown low-rank matrix $\boldsymbol{M} \in \mathbb{R}^{d_1 \times d_2}$ from an observed, yet incomplete, matrix. Let $\boldsymbol{X}$ denote a matrix whose entries are drawn from a normal mean–variance mixture distribution. We use μ_{ij} and x_{ij} to represent the (i, j)-th entries of $\boldsymbol{M}$ and $\boldsymbol{X}$, respectively. Based on the normal mean-variance model, the relationship between μ_{ij} and x_{ij} is given by

$$x_{ij} = \mu_{ij} + z\xi + z^{1/2}\sigma\epsilon,$$

where μ_{ij} serves as the location parameter of x_{ij} and both ξ and σ are given constants. In the quantization scenario, we have $\boldsymbol{Y} = \mathcal{Q}(\boldsymbol{X})$. The entire matrix $\boldsymbol{Y}$ is not available for observation. Let $\boldsymbol{\Omega}$ denote the index set of the observed entries. Our goal is to recover $\boldsymbol{M}$ from incomplete $\boldsymbol{Y}$, which is equivalent to estimating all the μ_{ij} in the matrix $\boldsymbol{M}$.

Since x_{ij} follows a univariate normal mean-variance model and $\boldsymbol{Y} = \mathcal{Q}(\boldsymbol{X})$, the density function of y_{ij} is identical with (5) under the univariate case. Hence, the negative log-likelihood function for y_{ij} with $ij \in \boldsymbol{\Omega}$ is

$$\sum_{(i,j)\in\boldsymbol{\Omega}} \log p\left(y_{ij} \mid \boldsymbol{M}\right),$$

which is the objective function of (16). Applying Jensen's inequality as in (9), we have the surrogate function

$$\begin{aligned} S(\boldsymbol{M}; \underline{\boldsymbol{M}}) = \sum_{(i,j)\in\boldsymbol{\Omega}} \Bigg[& \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}\left[\log p(z_t)\right] - \frac{1}{2}\log\det\sigma^2 \\ & - \frac{1}{2\sigma^2}\left(\left(u_{ij} - 2v_{ij}\mu + \iota_{ij}\mu^2\right) - 2(w_{ij} - \mu)\xi + \zeta_{ij}\xi^2\right)\Bigg] + \text{const.}, \end{aligned}$$

where u_{ij}, v_{ij}, and w_{ij} are the univariate versions of $\boldsymbol{U}_{ij}$, $\boldsymbol{v}_{ij}$, and $\boldsymbol{w}_{ij}$, respectively. Ignoring the terms which are independent with μ_{ij}, the surrogate function becomes

$$S(\boldsymbol{M}; \underline{\boldsymbol{M}}) = \frac{1}{2\sigma^2} \sum_{(i,j)\in\boldsymbol{\Omega}} \left(2v_t\mu - \iota_t\mu^2 - 2\mu\xi\right) + \text{const.}$$

Making a low-rank factorization to the matrix $\boldsymbol{M}$ as $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{B}^\top$, we have $\mu_{ij} = \boldsymbol{a}_i\boldsymbol{b}_j^\top$. Setting $e_{ij} = \frac{v_{ij}-\xi}{\iota_{ij}}$, we have that maximizing $S(\boldsymbol{M}; \underline{\boldsymbol{M}})$ is equivalent to maximize

$$\sum_{(i,j)\in\boldsymbol{\Omega}} \left(\boldsymbol{a}_i\boldsymbol{b}_j^\top - e_{ij}\right)^2,$$

which is identical with (17).

D.2 Surrogate Function Derivation in Quantized Compressive Sensing

In quantized compressive sensing, the base model with normal mean-variance mixture noise is given by

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{x}), \quad \boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\vartheta} + z\boldsymbol{\xi} + z^{1/2}\sigma\boldsymbol{\epsilon}.$$

In the above model, the term $\boldsymbol{\Phi}\boldsymbol{\vartheta}$ can be regarded as the location parameter of $\boldsymbol{x}$. We can use the ML estimation method to recover the sparse signal $\boldsymbol{\vartheta}$, which has the optimization problem

$$\max_{\boldsymbol{\vartheta}} \quad \log p\left(\boldsymbol{y} \mid \boldsymbol{\vartheta}\right).$$

Following the suggestions from Zymnis et al. (2009), we add a ℓ_1-regularization term to force the solution of $\boldsymbol{\vartheta}$ to be sparse. Hence, we obtain the optimization problem

$$\max_{\boldsymbol{\vartheta}} \quad \log p\left(\boldsymbol{y} \mid \boldsymbol{\vartheta}\right) + \eta\|\boldsymbol{\vartheta}\|_1.$$

To solve the above optimization problem through ECM algorithm, we can apply the E-step as in (9) to obtain the surrogate function

$$\begin{aligned} S(\boldsymbol{\vartheta}; \underline{\boldsymbol{\vartheta}}) = \mathsf{E}_{z_t|\boldsymbol{y}_t;\underline{\boldsymbol{\theta}}}\left[\log p(z_t)\right] - \frac{1}{2}\log\det\sigma^2 \\ - \frac{1}{2\sigma^2}\left(\left(\boldsymbol{u} - 2\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{\vartheta} + \iota\boldsymbol{\vartheta}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{\vartheta}\right) - 2(\boldsymbol{w} - \boldsymbol{A}\boldsymbol{\vartheta})^\top\boldsymbol{\xi} + \zeta\boldsymbol{\xi}^\top\boldsymbol{\xi}\right) + \eta\|\boldsymbol{\vartheta}\|_1 + \text{const}. \end{aligned}$$

Ignoring the terms which are independent with $\boldsymbol{\vartheta}$, the surrogate function becomes

$$S(\boldsymbol{\vartheta}; \underline{\boldsymbol{\vartheta}}) = \frac{1}{2\sigma^2}\left(2\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{\vartheta} - \iota\boldsymbol{\vartheta}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{\vartheta} - 2\boldsymbol{\vartheta}^\top\boldsymbol{A}^\top\boldsymbol{\xi}\right) + \eta\|\boldsymbol{\vartheta}\|_1 + \text{const}.$$

Setting $\boldsymbol{e} = \frac{\boldsymbol{v}-\boldsymbol{\xi}}{\iota}$, we have that maximizing $S(\boldsymbol{\vartheta}; \underline{\boldsymbol{\vartheta}})$ is equivalent to maximize

$$\|\boldsymbol{A}\boldsymbol{\vartheta} - \boldsymbol{e}\|_2^2 + \eta\left\|\boldsymbol{\vartheta}\right\|_1,$$

which is identical with the surrogate function in quantized compressive sensing.

E Experiment Details

E.1 Benchmark Settings

Quantized matrix completion: For all algorithms presented in Table 2, the complete matrix $\boldsymbol{M}$ is reconstructed from the training data. Denote by $\boldsymbol{\Omega}_{\text{test}}$ the set of indices corresponding to entries in the test data. The definitions of the two benchmarks are given as follows.

1. Accuracy: $\sum_{(i,j)\in\boldsymbol{\Omega}_{\text{test}}} I(\mathcal{Q}(\mu_{ij}) = y_{ij})$;
2. RMSE: $\sqrt{\frac{1}{n_{\text{test}}}\sum_{(i,j)\in\boldsymbol{\Omega}_{\text{test}}}(\mu_{ij} - y_{ij})^2}$,

where $I(\cdot)$ is the indicator function.

Quantized compressive sensing: The signal-to-noise ratio (SNR) is defined as the ratio of the expectation of the original measurements to the variance of the noise term, which is

$$\text{SNR} = \frac{\text{E}[\boldsymbol{A}\boldsymbol{\vartheta}]}{\text{Var}[\boldsymbol{x} - \boldsymbol{A}\boldsymbol{\vartheta}]}.$$

Denote the ground truth sparse signal as $\boldsymbol{\vartheta}_{\text{gd}}$, the estimated sparse signal as $\boldsymbol{\vartheta}_{\text{est}}$. The cosine similarity is defined as

$$\text{Cos Sim} = \frac{\boldsymbol{\vartheta}_{\text{gd}}^\top\boldsymbol{\vartheta}_{\text{est}}}{\|\boldsymbol{\vartheta}_{\text{gd}}\|_2\|\boldsymbol{\vartheta}_{\text{est}}\|_2}.$$

E.2 Experimental Settings

In Figure 4, we set $\mu_i \sim \text{Uniform}(0,1)$, $\xi_i \sim \text{Uniform}(0,2)$, and $\mathbf{v}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ for $i = 1, \ldots, d$, and construct the covariance matrix as $\boldsymbol{\Sigma} = \sum_{i=1}^{d} \mathbf{v}_i \mathbf{v}_i^\top$. For the GH distribution, we use $\lambda = -0.5$, $\delta = 2$, and $\gamma = 2$; for GHST, we set $\nu = 5$, $\lambda = -\mu/2$, $\delta = \nu$, and $\gamma = 0$. The number of samples is $n = 1 \times 10^4$ and the threshold is fixed at $\tau = 0.3$.

The specific parameter settings for the algorithms in Table 2 are given by

1. Standard Gaussian: $r = 2$ and $\sigma^2 = 0.8$;
2. one-bit Gaussian: $r = 2$ and $\sigma^2 = 0.8$;
3. multi-bit Gaussian: $r = 2$ and $\sigma^2 = 0.8$;
4. multi-bit Student's t: $r = 2$, $\sigma^2 = 0.8$, and $\nu = 6$;
5. multi-bit GHST: $r = 2$, $\sigma^2 = 0.8$, $\xi = -0.1$, and $\nu = 8$;
6. multi-bit GH: $r = 2$, $\sigma^2 = 0.8$, $\xi = -0.1$, $\lambda = -4$, $\delta = 4$, and $\gamma = 0.3$.