Efficient CRB Estimation for Linear Models via Expectation Propagation and Monte Carlo Sampling

Fangqing Xiao, *Member, IEEE*, and Dirk Slock, *Life Fellow, IEEE*

***Abstract*—The Cramér–Rao bound (CRB) quantifies the variance lower bound for unbiased estimators, but it is intractable to evaluate in linear hierarchical Bayesian models with non-Gaussian priors due to the intractable marginal likelihood. Existing methods, including variational Bayes and Markov chain Monte Carlo (MCMC)-based approaches, often have high computational cost and slow convergence. We propose an efficient framework to approximate the Fisher information matrix (FIM) and the CRB by expressing the gradient of the log marginal likelihood as a posterior expectation. Expectation propagation (EP) is used to approximate the posterior as a Gaussian, enabling accurate moment estimation compared to pure sampling-based methods. Numerical experiments on small-scale sparse models show that the EP-based CRB approximation achieves lower average normalized mean squared error (NMSE) and faster convergence than classical baselines in non-Gaussian settings.**

***Index Terms*—Cramér–Rao bound, expectation propagation, fisher information matrix, Monte Carlo sampling.**

I. Introduction

THE Cramér–Rao bound (CRB) provides a fundamental benchmark for the performance of unbiased estimators, establishing a theoretical lower bound on the estimation variance [1], [2]. In linear hierarchical Bayesian models with non-Gaussian priors, direct evaluation of the CRB is often intractable due to the complexity of the marginal likelihood and its derivatives.

Traditional approaches to approximating the CRB, such as the Laplace approximation [3], variational Bayes (VB) inference [4], and fully numerical methods based on Markov chain Monte Carlo (MCMC) sampling [5], can approximate posterior covariances or Hessians but typically suffer from high computational cost or limited accuracy when relying on a single global Gaussian/mean-field approximation under strongly non-Gaussian posteriors. Direct Monte Carlo sampling of posterior gradients or Fisher information matrix (FIM) terms also faces challenges, including slow convergence in scenarios with limited sample sizes [6].

This work was supported in part by the industrial members such as ORANGE, BMW, SAP, iABG, and Norton LifeLock, in part by Franco-German Project 5G-OPERA (BPI), in part by the French Project YACARI (PEPR-5G), in part by EU INFRA Project CONVERGE, and in part by Huawei France funded Chair towards Future Wireless Networks.

The authors are with EURECOM, F-06904 Sophia Antipolis, France (e-mail: fangqing.xiao@eurecom.fr; dirk.slock@eurecom.fr).

To address these limitations, we propose a flexible and efficient framework for approximating the FIM and thus the CRB in complex hierarchical models. The key idea is to express the gradient of the log marginal likelihood as an expectation over the posterior distribution of latent variables, transforming an intractable marginal gradient computation into a tractable posterior expectation. However, direct sampling-based approximations of these posterior expectations can be computationally demanding [7].

Instead, we leverage adaptive expectation propagation (EP) [8], [9], [10], [11] to obtain a Gaussian-family approximation via iterative tilted moment matching, enabling tractable posterior moments while accommodating non-Gaussian prior factors. This approach first generates a limited number of marginal samples from the generative model and, for each sample, applies EP to approximate the corresponding posterior distribution. Because EP directly yields the posterior means and variances, only a modest number of marginal samples is needed to accurately estimate the FIM. Compared to direct Monte Carlo methods [7] and message-passing methods like approximate message passing (AMP) [12], vector AMP (VAMP) [13], and generalized AMP (GAMP), our EP-based approach achieves significantly higher precision and faster convergence by directly leveraging posterior moment matching—particularly important for non-Gaussian hierarchical models.

The remainder of this letter is organized as follows. Section II formulates the problem and presents the CRB expression. Section III introduces the two-step approximation framework for the FIM. Section IV details the EP inference algorithm. Section V discusses implementation and algorithmic considerations. Section VI presents numerical experiments validating the proposed approach, and Section VII concludes this letter.

II. Problem Formulation and CRB Expression

We consider the following linear observation model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ denotes the observed vector, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a known observation matrix, $\mathbf{x} \in \mathbb{R}^N$ is the unknown signal vector (latent variable vector), and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ represents Gaussian noise with unknown variance σ^2.

The entries of $\mathbf{x}$ are assumed to be independent and identically distributed (i.i.d.) according to a prior distribution $p(x_n; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^P$ denotes a vector of hyperparameters characterizing the prior. Specifically, the prior distribution of $\mathbf{x}$ can be

written as

$$p(\mathbf{x};\boldsymbol{\theta}) = \prod_{n=1}^{N} p(x_n;\boldsymbol{\theta}). \tag{2}$$

For mathematical convenience, we adopt the i.i.d. assumption in this work. Nevertheless, our framework can be extended to the more general case in which the entries of $\mathbf{x}$ are independent but not identically distributed (non-i.i.d.). In such cases, only minor modifications are required in the subsequent derivations, and the overall approach remains applicable.

Our primary objective is to jointly estimate the hyperparameters $\boldsymbol{\theta}$ of the prior distribution and the noise variance σ^2. To quantify the fundamental lower bound on the variance of any unbiased estimator $\hat{\boldsymbol{\phi}}$ of the unknown hyperparameter vector $\boldsymbol{\phi} = [\boldsymbol{\theta}^\top, \sigma^2]^\top$, we employ the *Cramér-Rao Bound (CRB)*. The CRB states that the covariance matrix of any unbiased estimator satisfies

$$\mathrm{Cov}(\hat{\boldsymbol{\phi}}) \succeq \mathrm{CRB}(\boldsymbol{\phi}) = \mathbf{J}^{-1}(\boldsymbol{\phi}), \tag{3}$$

where the *Fisher Information Matrix (FIM)* is defined by

$$\mathbf{J}(\boldsymbol{\phi}) = \mathbb{E}_{\mathbf{y}|\boldsymbol{\phi}}\left[\nabla_{\boldsymbol{\phi}} \log p(\mathbf{y};\boldsymbol{\phi}) \nabla_{\boldsymbol{\phi}}^\top \log p(\mathbf{y};\boldsymbol{\phi})\right]. \tag{4}$$

As in (4), the marginal likelihood can be expressed as

$$p(\mathbf{y}|\boldsymbol{\phi}) = \int p(\mathbf{y}|\mathbf{x},\sigma^2)\, p(\mathbf{x};\boldsymbol{\theta})\, d\mathbf{x}, \tag{5}$$

which is generally intractable due to the high-dimensional integral. To address this challenge, we leverage the standard derivation in the expectation-maximization literature [14]:

$$\nabla_{\boldsymbol{\phi}} \log p(\mathbf{y};\boldsymbol{\phi}) = \mathbb{E}_{p(\mathbf{x}|\mathbf{y};\boldsymbol{\phi})}\left[\nabla_{\boldsymbol{\phi}} \log p(\mathbf{y},\mathbf{x};\boldsymbol{\phi})\right]. \tag{6}$$

Define the complete-data score function as following:

$$\mathbf{B}(\mathbf{x},\mathbf{y};\boldsymbol{\phi}) = \nabla_{\boldsymbol{\phi}} \log p(\mathbf{y},\mathbf{x};\boldsymbol{\phi}), \tag{7}$$

which can be decomposed as

$$\mathbf{B}(\mathbf{x},\mathbf{y};\boldsymbol{\phi}) = \nabla_{\boldsymbol{\phi}} \log p(\mathbf{y}|\mathbf{x},\sigma^2) + \nabla_{\boldsymbol{\phi}} \log p(\mathbf{x};\boldsymbol{\theta}). \tag{8}$$

This explicit form highlights the tractability of evaluating $\mathbf{B}(\mathbf{x},\mathbf{y};\boldsymbol{\phi})$ once the posterior distribution is available.

Substituting the marginal score expression (6) into (4), the FIM becomes

$$\mathbf{J}(\boldsymbol{\phi}) = \mathbb{E}_{\mathbf{y};\boldsymbol{\phi}}\Big[\mathbb{E}_{p(\mathbf{x}|\mathbf{y};\boldsymbol{\phi})}\left[\mathbf{B}(\mathbf{x},\mathbf{y};\boldsymbol{\phi})\right] \times \mathbb{E}_{p(\mathbf{x}|\mathbf{y};\boldsymbol{\phi})}\left[\mathbf{B}(\mathbf{x},\mathbf{y};\boldsymbol{\phi})\right]^\top\Big], \tag{9}$$

which directly connects the FIM to the outer product of the posterior-averaged complete-data score.

III. Fisher Information Matrix Approximation

We exploit the factorization structure in (9) to decompose the joint expectation into nested conditional expectations:

$$\mathbf{J}(\boldsymbol{\phi}) = \mathbb{E}_{p(\mathbf{y};\boldsymbol{\phi})}\left[\mathbf{m}(\mathbf{y})\mathbf{m}(\mathbf{y})^\top\right], \mathbf{m}(\mathbf{y}) = \mathbb{E}_{p(\mathbf{x}|\mathbf{y},\boldsymbol{\phi})}[\mathbf{B}(\mathbf{x},\mathbf{y};\boldsymbol{\phi})]. \tag{10}$$

This separation naturally suggests a *two-step approximation* scheme:

i) *Outer Monte Carlo Sampling:* When possible, generate samples $\{\mathbf{y}^{(s)}\}_{s=1}^{S}$ from the marginal distribution $p(\mathbf{y};\boldsymbol{\phi})$ by simulating from the generative model $p(\mathbf{y}|\mathbf{x};\boldsymbol{\phi})$ and the prior $p(\mathbf{x};\boldsymbol{\theta})$. Alternatively, use available observed data $\{\mathbf{y}^{(s)}\}$ when the marginal model is intractable.

ii) *Inner Posterior Expectation:* For each $\mathbf{y}^{(s)}$, approximate $\mathbf{m}(\mathbf{y}^{(s)})$ by computing the posterior expectation under $p(\mathbf{x}|\mathbf{y}^{(s)};\boldsymbol{\phi})$.

This factorized *Monte Carlo + posterior expectation* framework circumvents explicit posterior expectation over $p(\mathbf{x}|\mathbf{y};\boldsymbol{\phi})$.

In the outer Monte Carlo step, we approximate the expectation over $\mathbf{y}$ by drawing S samples $\{\mathbf{y}^{(s)}\}_{s=1}^{S}$ from the marginal distribution $p(\mathbf{y};\boldsymbol{\phi})$, which is generally intractable due to the integral over the latent variable $\mathbf{x}$ in (5). When the generative model is available, samples of $\mathbf{y}$ can be obtained by first drawing $\mathbf{x}^{(s)} \sim p(\mathbf{x};\boldsymbol{\theta})$ from the prior and then computing $\mathbf{y}^{(s)} = \mathbf{A}\mathbf{x}^{(s)} + \mathbf{v}^{(s)}$, where $\mathbf{v}^{(s)} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ represents measurement noise. This simulation-based approach directly reflects the generative process implied by the model and ensures that the samples $\mathbf{y}^{(s)}$ are consistent with the underlying system parameters.

In practical scenarios where the generative model is unavailable or $p(\mathbf{y};\boldsymbol{\phi})$ cannot be evaluated, available observed data $\{\mathbf{y}^{(s)}\}$ can be used directly. This bypasses the need to compute the marginal distribution explicitly and leverages real-world measurements for performance evaluation. The sample size S balances computational cost and approximation accuracy: larger S improves the outer expectation approximation but at increased computational expense. Each sampled $\mathbf{y}^{(s)}$ subsequently serves as input to the inner posterior expectation approximation described in the following section.

IV. Expectation Propagation Inference

Given the sampled $\mathbf{y}$, we consider the intractable posterior

$$p(\mathbf{x}|\mathbf{y};\boldsymbol{\phi}) = \frac{p(\mathbf{y}|\mathbf{x};\sigma^2)\prod_{n=1}^{N} p(x_n;\boldsymbol{\theta})}{\int p(\mathbf{y}|\mathbf{x};\sigma^2)\prod_{n=1}^{N} p(x_n;\boldsymbol{\theta}) d\mathbf{x}}, \tag{11}$$

where $\boldsymbol{\phi} = [\boldsymbol{\theta}^\top, \sigma^2]^\top$. The prior factors $p(x_n;\boldsymbol{\theta})$ are typically non-Gaussian, rendering exact inference infeasible.

To address this, we employ Expectation Propagation (EP) [8], [9], which directly approximates each prior factor by a Gaussian:

$$f_n(x_n) = \mathcal{N}(x_n; p_n, \tau_n), \tag{12}$$

yielding the approximate posterior

$$q(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x};\sigma^2)\prod_{n=1}^{N} f_n(x_n)}{\int p(\mathbf{y}|\mathbf{x};\sigma^2)\prod_{n=1}^{N} f_n(x_n) d\mathbf{x}}, \tag{13}$$

where $q(\mathbf{x}|\mathbf{y})$ are the approximate Gaussian posterior with mean $\mathbf{m}$ and covariance $\mathbf{C}_{\mathbf{m}}$.

A. Initialization

We initialize the Gaussian factors $f_n(x_n)$ by matching the first and second moments of the true prior:

$$p_n = \mathbb{E}[x_n], \quad \tau_n = \mathrm{Var}[x_n], \tag{14}$$

so that the initial $q(\mathbf{x}|\mathbf{y})$ corresponds to the linear minimum mean square error (LMMSE) estimate. This initialization provides a practical starting point, especially when no other information is available.

B. Iterative EP Updates

The EP algorithm refines the Gaussian factors iteratively as follows:

1) Posterior Approximation: Given the current Gaussian factors, the approximate posterior is updated by

$$q(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x};\sigma^2)\prod_{n=1}^{N} f_n(x_n)}{\int p(\mathbf{y}|\mathbf{x};\sigma^2)\prod_{n=1}^{N} f_n(x_n)\,d\mathbf{x}}, \tag{15}$$

where $q(\mathbf{x}|\mathbf{y})$ remains Gaussian with mean $\mathbf{m}$ and covariance $\mathbf{C_m}$. These parameters can be computed in closed form:

$$\mathbf{C_m} = \left(\sigma^{-2}\mathbf{A}^T\mathbf{A} + \mathbf{C_p}^{-1}\right)^{-1}, \tag{16}$$

$$\mathbf{m} = \mathbf{C_m}\left(\sigma^{-2}\mathbf{A}^T\mathbf{y} + \mathbf{C_p}^{-1}\mathbf{p}\right), \tag{17}$$

where $\mathbf{p}$ collects the means p_n of the approximate prior factors, and $\mathbf{C_p}$ is a diagonal matrix with the variances τ_n as its diagonal entries. Here, $(\cdot)^T$ denotes the matrix transpose.

2) Extrinsic Distribution: For each variable x_n, the extrinsic distribution is computed by removing $f_n(x_n)$ from $q(\mathbf{x}|\mathbf{y})$:

$$b(x_n) = \int \frac{q(\mathbf{x}|\mathbf{y})}{f_n(x_n)}\,d\mathbf{x}_{\setminus n}. \tag{18}$$

Because $q(\mathbf{x}|\mathbf{y})$ is Gaussian, this marginalization is tractable, yielding a Gaussian $b(x_n)$ that captures the information about x_n from all other factors.

3) Tilted Distribution and Moment Matching: The *tilted distribution* combines $b(x_n)$ with the exact prior:

$$\tilde{p}(x_n) \propto b(x_n)p(x_n;\boldsymbol{\theta}). \tag{19}$$

Since $\tilde{p}(x_n)$ is generally non-Gaussian, it is approximated by matching its first and second moments:

$$q(x_n) = \text{proj}[\tilde{p}(x_n)], \tag{20}$$

where proj[] denotes the Gaussian projection that matches the first two moments. The new factor update is

$$f_n^{\text{new}}(x_n) \propto \frac{q(x_n)}{b(x_n)}. \tag{21}$$

4) Damping for Stability: To ensure robust convergence and prevent oscillations, a damping step is applied:

$$f_n(x_n) \leftarrow f_n(x_n)^{1-\eta}\cdot f_n^{\text{new}}(x_n)^{\eta}, \quad \eta \in (0,1]. \tag{22}$$

A typical choice is $\eta = 0.5$, which balances convergence speed and stability. In practice, η can also be adapted during iterations based on the magnitude of factor updates, reducing η when large oscillations are detected and increasing it as the algorithm stabilizes. This adaptive damping can further improve convergence in challenging scenarios. These steps are repeated for all n until convergence. If an update yields non-positive values, we apply damping and project them to a small positive threshold.

5) Convergence: The iterative updates are repeated until convergence is reached. A typical convergence criterion is to monitor the change in the means and variances of the Gaussian factors $f_n(x_n)$ across iterations. Convergence is declared when the maximum relative change across all n falls below a predefined threshold (e.g., 10^{-3}). This ensures that the approximate posterior $q(\mathbf{x}|\mathbf{y})$ has stabilized and no longer undergoes significant updates.

This EP framework systematically replaces intractable expectations with tractable Gaussian updates. Initialization with prior moments ensures a sensible starting point, while damping promotes convergence stability.

Algorithm 1: Proposed Algorithm for Fisher Information Matrix Approximation.

Require: Number of outer samples S, observation matrix $\mathbf{A}$, prior parameters $\boldsymbol{\theta}$, noise variance σ^2, damping factor η

Ensure: Approximation of the Fisher Information Matrix $\mathbf{J}(\boldsymbol{\phi})$

1: **for** $s = 1, \ldots, S$ **do**
2: Generate latent signal $\mathbf{x}^{(s)} \sim p(\mathbf{x};\boldsymbol{\theta})$
3: Generate observation $\mathbf{y}^{(s)} = \mathbf{A}\mathbf{x}^{(s)} + \mathbf{v}^{(s)}$, with $\mathbf{v}^{(s)} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M)$
4: Approximate the posterior $p(\mathbf{x}|\mathbf{y}^{(s)};\boldsymbol{\phi})$ by EP:

$$p(\mathbf{x}|\mathbf{y}^{(s)};\boldsymbol{\phi}) \approx q(\mathbf{x}|\mathbf{y}^{(s)}) = \mathcal{N}(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})$$

5: Compute the approximate posterior gradient (posterior average of complete-data score):

$$\mathbf{m}(\mathbf{y}^{(s)}) = \mathbb{E}_{q(\mathbf{x}|\mathbf{y}^{(s)})}[\nabla_{\boldsymbol{\phi}} \log p(\mathbf{y}^{(s)}, \mathbf{x};\boldsymbol{\phi})].$$

6: **end for**
7: Approximate the Fisher Information Matrix:

$$\mathbf{J}(\boldsymbol{\phi}) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbf{m}(\mathbf{y}^{(s)})\mathbf{m}(\mathbf{y}^{(s)})^{\top}.$$

V. Algorithm Implementation Details

Algorithm 1 summarizes the entire procedure for approximating the Fisher Information Matrix (FIM) by combining:

- *Outer Monte Carlo sampling of* $\mathbf{y}$ ($s = 1, \ldots, S$): captures variability of $\mathbf{y}$ in the marginal FIM.
- *Inner EP-based Gaussian approximation of* $p(\mathbf{x}|\mathbf{y})$: directly provides a posterior mean and covariance.
- *Direct Gaussian moment calculation for posterior expectations* $\mathbf{m}(\mathbf{y})$: no further sampling is needed.

This approach avoids high-dimensional joint sampling of $(\mathbf{x}, \mathbf{y})$, replacing it with a two-level hierarchical approximation. It leverages the analytic tractability of Gaussian expectations while retaining the flexibility of EP to adapt to non-Gaussian prior structures (e.g., spike-and-slab). In practice, the number of outer samples S is chosen based on convergence diagnostics and desired accuracy, while the EP typically converges in a few iterations per observation.

This combination of adaptive moment-matching and outer Monte Carlo sampling provides an accurate, scalable, and numerically robust way to approximate the FIM in complex latent variable models.

VI. Experimental Evaluation

A. Setup and Exact CRB Computation

We consider a spike-and-slab prior for the unknown vector $\mathbf{x} \in \mathbb{R}^N$:

$$p(x_n) = \pi\,\delta(x_n) + (1-\pi)\,\mathcal{N}(x_n; 0, \tau^2), \tag{23}$$

with $\pi = 0.5$, $\tau^2 = 1$, and observation noise variance $\sigma^2 = 1$. The measurement matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ is generated with i.i.d. Gaussian entries and normalized column-wise. We focus on $N = 10$, $M = 5$, where exact posterior computations are tractable.

For each observation $\mathbf{y}$ and sparse signal $\mathbf{x}$, the complete-data gradient vector $\mathbf{B}(\mathbf{x}, \mathbf{y}; \boldsymbol{\phi})$ is given by

$$\mathbf{B}(\mathbf{x}, \mathbf{y}; \boldsymbol{\phi}) = \begin{bmatrix} \sum_{n=1}^{N} \frac{\partial}{\partial \tau^2} \log p(x_n \mid \tau^2) \\ [1ex] - \frac{M}{2\sigma^2} + \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{2\sigma^4} \end{bmatrix}. \tag{24}$$

Here, the derivative w.r.t. τ^2 depends only on $\mathbf{x}^2$ terms and can be written as:

$$\frac{\partial}{\partial \tau^2} \log p(x_n \mid \tau^2) = \begin{cases} -\frac{1}{2\tau^2} + \frac{x_n^2}{2\tau^4}, & x_n \neq 0, \\ [1ex] \frac{-\frac{1-\pi}{2\tau^2} \frac{1}{\sqrt{2\pi\tau^2}}}{\pi + \frac{1-\pi}{\sqrt{2\pi\tau^2}}}, & x_n = 0. \end{cases} \tag{25}$$

Consequently, computing the exact CRB requires only the posterior first and second moments of $\mathbf{x}$. In this low-dimensional setting, the exact posterior average

$$\mathbf{m}_{\text{exact}}(\mathbf{y}) = \mathbb{E}_{p(\mathbf{x}|\mathbf{y};\boldsymbol{\phi})}[\mathbf{B}(\mathbf{x}, \mathbf{y}; \boldsymbol{\phi})] \tag{26}$$

can be computed in closed form by enumerating all posterior configurations, as detailed in [15]. Although this approach provides accurate CRB evaluation in small-scale problems, it is computationally prohibitive for larger dimensions. The exact FIM and CRB are then given by

$$\mathbf{J}_{\text{exact}} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{m}_{\text{exact}}(\mathbf{y}^{(s)}) \mathbf{m}_{\text{exact}}(\mathbf{y}^{(s)})^{\top}, \tag{27}$$

$$\text{CRB}_{\text{exact}} = \mathbf{J}_{\text{exact}}^{-1}. \tag{28}$$

B. Approximation Methods and Evaluation

For benchmarking, we generate S independent observations $\mathbf{y}^{(s)}$. Since the complete-data gradient expectations $\mathbf{m}(\mathbf{y})$ depend only on posterior first and second moments of $\mathbf{x}$, the approximate methods are designed to estimate these moments efficiently:

- *EP-based (proposed):* Adaptive-damped EP approximates the posterior as Gaussian, directly yielding the required first and second moments.
- *AMP:* Uses approximate message passing to compute approximate posterior means and variances.
- *LMMSE baseline:* Assumes Gaussian prior matching prior moments, yielding analytical approximate posterior moments.
- *Blocked Gibbs MCMC [7], [16]:* Employs the blocked Gibbs sampling MCMC scheme from [16] to replace the basic MCMC in [7], using $L = 100, 500, 2500$ inner iterations to generate posterior samples and compute posterior means and variances.

Each method estimates the approximate posterior gradient $\mathbf{m}_{\text{approx}}(\mathbf{y}^{(s)})$ and constructs

$$\mathbf{J}_{\text{approx}} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{m}_{\text{approx}}(\mathbf{y}^{(s)}) \mathbf{m}_{\text{approx}}(\mathbf{y}^{(s)})^{\top}, \tag{29}$$

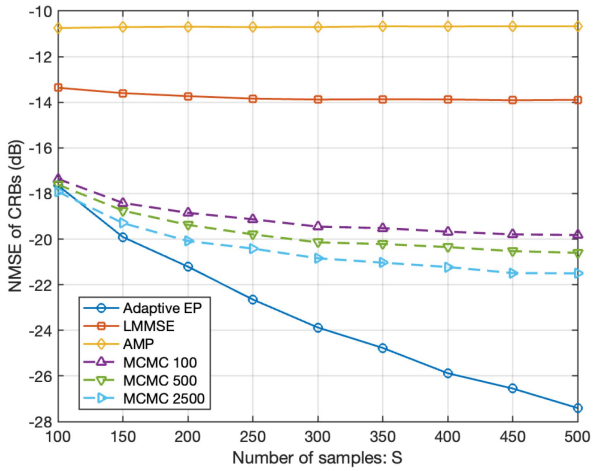

Fig. 1. Average NMSE of the estimated CRB diagonal variance bounds compared to the exact CRB across 5000 independent repetitions.

$$\text{CRB}_{\text{approx}} = \mathbf{J}_{\text{approx}}^{-1}. \tag{30}$$

Convergence is quantified by the normalized mean squared error (NMSE) of the CRB diagonal elements (variance bounds):

$$\text{NMSE} = \frac{1}{(P+1)} \sum_{n=1}^{P+1} \frac{(\hat{d}_n - d_n)^2}{d_n^2}, \tag{31}$$

where $\hat{d}_n$ and d_n denote the n-th diagonal entries of $\text{CRB}_{\text{approx}}$ and $\text{CRB}_{\text{exact}}$, respectively. We perform 5000 repetitions for each S, varying from 100 to 500 in steps of 50.

C. Results

Fig. 1 shows the average NMSE convergence of the CRB diagonal variance bounds compared to the exact benchmark. The EP-based method achieves the best convergence and lowest average NMSE, improving steadily with S. MCMC-based estimates converge more slowly but benefit from larger L. The LMMSE baseline outperforms AMP, which exhibits the slowest convergence and largest average NMSE due to mean-field limitations in this small dimension setting.

VII. Conclusion

We have presented a scalable framework for approximating the Fisher information matrix and the CRB in linear hierarchical Bayesian models with non-Gaussian priors. By expressing the intractable marginal gradient as a posterior expectation and leveraging expectation propagation to approximate the posterior, our method achieves accurate and computationally efficient FIM estimation. Numerical experiments demonstrate that the proposed EP-based approach outperforms traditional sampling-based methods, particularly in scenarios with strongly non-Gaussian priors. This highlights the potential of combining EP-based approximations with Monte Carlo outer sampling for robust and scalable CRB estimation in high-dimensional signal processing applications. Future work will focus on extending this framework to large-scale models, paving the way for practical deployment in real-world signal processing and wireless communication systems.

REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[2] F. Nielsen, "Cramér–Rao lower bound and information geometry," in *Connected at Infinity II, Texts and Readings in Mathematics*. R. Bhatia, C. S. Rajan, and A. I. Singh, Eds., Gurgaon, India: Hindustan Book Agency, 2013, pp. 18–37, doi: 10.1007/978-93-86279-56-9_2.

[3] L. Tierney and J. B. Kadane, "Accurate approximations for posterior moments and marginal densities," *J. Amer. Statist. Assoc.*, vol. 81, no. 393, pp. 82–86, 1986.

[4] H. Attias, "A variational Bayesian framework for graphical models," *Neural Comput.*, vol. 12, no. 4, pp. 909–932, 2000.

[5] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer, 2004.

[6] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér–Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[7] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Computing the Cramer–Rao bound of Markov random field parameters: Application to the Ising and the Potts models," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 47–50, Jan. 2014.

[8] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.*, San Francisco, CA, USA, 2001, pp. 362–369.

[9] Y. Li and J. M. Hernández-Lobato, and R. E. Turner, "Stochastic expectation propagation," in *Proc. 29th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 2323–2331.

[10] G. Dehaene and S. Barthelmé, "Expectation propagation in the large data limit," *J. Roy. Statist. Soc. Ser. B Methodol.*, vol. 80, no. 1, pp. 199–217, Jan. 2018.

[11] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory*, vol. 67, no. 3, pp. 1544–1571, Jan. 2020.

[12] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.

[13] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.: Ser. B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[15] Z. Zhao, F. Xiao, and D. Slock, "Approximate message passing for not so large niid generalized linear models," in *Proc. IEEE 24th Int. Workshop Signal Process. Adv. Wireless Commun.*, 2023, pp. 386–390.

[16] M. Amrouche, H. Carfantan, and J. Idier, "Efficient Sampling of Bernoulli-Gaussian-mixtures for sparse signal restoration," *IEEE Trans. Signal Process.*, vol. 70, pp. 5578–5591, 2022.