Sigma-Point Expectation Propagation

Fangqing Xiao, Zunqi Li, Dirk Slock
EURECOM, Sophia Antipolis, France
Email: {fangqing.xiao, zunqi.li, dirk.slock}@eurecom.fr

***Abstract*—Nonlinear Bayesian inference problems often involve black-box forward models with strong curvature, for which Jacobians are inconvenient or unavailable. At the same time, priors are frequently non-Gaussian (e.g., Laplace or Student-t) to promote sparsity and robustness. Expectation propagation (EP) can in principle handle such priors, but in nonlinear settings it typically relies on first-order Taylor linearization of the forward map to obtain likelihood moments, leading to biased approximations at low signal-to-noise ratio (SNR) and coupling the algorithm to hand-crafted Jacobians. We propose sigma-point expectation propagation (SP-EP), a Jacobian-free EP framework for nonlinear, non-Gaussian inference. SP-EP embeds a sigma-point rule, instantiated here by the unscented transform, into the EP updates: for each latent variable we build an extrinsic likelihood by propagating a Gaussian cavity distribution over the remaining variables through the nonlinear measurement, then perform one-dimensional moment matching under the true prior. This yields a plug-and-play likelihood–moment module that leaves the EP shell unchanged. We further provide a simple damping and covariance-jitter recipe that ensures robustness. Simulations on a nonlinear three-dimensional model with Laplace and Student-t priors show that SP-EP consistently improves posterior-mean RMSE over a Taylor-based EP baseline when benchmarked against a Metropolis–Hastings MMSE reference across a range of SNRs.**

I. Introduction

We consider nonlinear sensing problems of the form

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{v}, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is a latent state, $\boldsymbol{y} \in \mathbb{R}^m$ is an observation, $f(\cdot)$ is a nonlinear (often black-box) map, and $\boldsymbol{v}$ is additive noise. In many applications $f(\cdot)$ is highly curved or only piecewise smooth, and analytic Jacobians are either unavailable or expensive to maintain. At the same time, non-Gaussian priors such as Laplace and Student-*t* distributions are routinely used to promote sparsity and robustness [2], [3], [10]. Practical Bayesian inference should therefore handle nonlinear measurements and heavy-tailed priors while treating $f(\cdot)$ as a black box.

Expectation propagation (EP) [1] is a flexible framework for approximate Bayesian inference with non-Gaussian priors. It maintains a Gaussian approximation to the posterior and iteratively updates local site factors by moment matching with respect to tilted distributions. EP and its variants have been successfully applied to sparse linear models and generalized linear models [8], [9], but for nonlinear models with $\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{v}$ the required likelihood moments under Gaussian cavities are intractable. A common workaround is to linearize $f(\cdot)$ around the current estimate and then run Gaussian EP on the resulting approximate model, in the spirit of the extended Kalman filter (EKF). Such Taylor-based schemes inherit the usual EKF limitations: they are biased under strong curvature, sensitive to the expansion point, and tied to hand-crafted Jacobians [7].

Sigma-point methods such as the unscented transform (UT) and cubature Kalman filtering (CKF) [4], [5] provide a Jacobian-free alternative. They propagate Gaussian uncertainty through nonlinear maps using deterministically chosen sigma points and accurately approximate means and covariances for moderately nonlinear transformations. Sigma-point Kalman filters and smoothers are widely used in nonlinear state estimation, but are typically restricted to Gaussian priors and do not exploit the site structure of EP. On the other hand, Markov chain Monte Carlo (MCMC) methods can handle general nonlinear/non-Gaussian posteriors but are often too expensive for repeated use in embedded or real-time systems.

Existing approaches to nonlinear, non-Gaussian inference thus fall into three categories: Taylor-based EP (non-Gaussian priors but Jacobian-dependent), sigma-point filtering/smoothing (Jacobian-free but essentially Gaussian priors), and MCMC (flexible but costly). To the best of our knowledge, combining sigma-point rules with EP in order to obtain a Jacobian-free, curvature-aware EP framework that natively supports heavy-tailed priors has not been explored.

A. Contributions

This paper proposes a *sigma-point expectation propagation* (SP-EP) framework that combines EP with sigma-point integration for nonlinear, non-Gaussian inference. Our main contributions are:

- We introduce a Jacobian-free likelihood–moment module inside EP based on the unscented transform. The EP "shell" remains unchanged, and non-Gaussian priors such as Laplace and Student-*t* are handled through standard EP sites.
- We give a practical implementation recipe including damping, covariance jitter, and cavity scaling, which yields robust convergence and keeps all covariance matrices symmetric positive definite in challenging regimes.
- We define a pluggable sigma-point interface that can be instantiated by UT, CKF, or other rules, letting practitioners trade off accuracy and cost without modifying the EP updates.
- On a nonlinear three-dimensional example with Laplace and Student-*t* priors, SP-EP consistently outperforms a Taylor-based EP baseline in posterior mean RMSE with

respect to a Metropolis–Hastings MMSE reference across a range of SNRs.

II. Problem Setup and EP Background

A. Nonlinear Sensing Model

The nonlinear forward map $f : \mathbb{R}^n \to \mathbb{R}^m$ in (1) is treated as a black-box routine, for which we can evaluate $f(\boldsymbol{x})$ but do not assume access to its Jacobian. The measurement noise $\boldsymbol{v}$ is modeled as

$$\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}), \tag{2}$$

with covariance $\boldsymbol{R} \in \mathbb{R}^{m\times m}$ positive definite. In our experiments, we take $\boldsymbol{R} = \sigma_v^2 \boldsymbol{I}_m$ and sweep the SNR by varying σ_v^2.

The prior on $\boldsymbol{x}$ is assumed separable and non-Gaussian:

$$p(\boldsymbol{x}) = \prod_{i=1}^{n} p_i(x_i). \tag{3}$$

Our goal is to approximate the posterior

$$p(\boldsymbol{x} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{x}) p(\boldsymbol{x}), \tag{4}$$

and in particular to estimate its mean and covariance, as well as quantify uncertainty, in a computationally efficient and robust way.

B. Expectation Propagation with Non-Gaussian Priors

EP [1] approximates a target distribution

$$p(\boldsymbol{x} \mid \boldsymbol{y}) \propto \prod_{\ell} t_\ell(\boldsymbol{x}) \tag{5}$$

by a tractable distribution $q(\boldsymbol{x})$, typically Gaussian, which factorizes as

$$q(\boldsymbol{x}) \propto \prod_{\ell} \tilde{t}_\ell(\boldsymbol{x}), \tag{6}$$

where each site $\tilde{t}_\ell(\boldsymbol{x})$ belongs to an exponential family. For Gaussian EP, $q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ and each site is Gaussian in (a subset of) the variables.

EP proceeds by iteratively updating sites:

1) Form the *cavity* distribution by removing the current site:

$$q_{-\ell}(\boldsymbol{x}) \propto \frac{q(\boldsymbol{x})}{\tilde{t}_\ell(\boldsymbol{x})}. \tag{7}$$

2) Form the *tilted* distribution

$$\hat{p}_\ell(\boldsymbol{x}) \propto q_{-\ell}(\boldsymbol{x})\, t_\ell(\boldsymbol{x}). \tag{8}$$

3) Fit a new Gaussian site $\tilde{t}_\ell^{\text{new}}(\boldsymbol{x})$ such that the moments of $q_{-\ell}(\boldsymbol{x})\tilde{t}_\ell^{\text{new}}(\boldsymbol{x})$ match those of $\hat{p}_\ell(\boldsymbol{x})$.

In practice, one often works with natural parameters and uses damping to stabilize the updates. When the prior is non-Gaussian but separable, each prior factor $p_i(x_i)$ is represented by a (possibly non-Gaussian) site that is approximated by a Gaussian factor in $q(\boldsymbol{x})$ via moment matching.

In linear or generalized linear models, moment matching is tractable or can be approximated efficiently. In nonlinear models with $\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{v}$, however, the tilted distribution involves the nonlinear likelihood $p(\boldsymbol{y} \mid \boldsymbol{x})$ and becomes intractable. Classical EP implementations linearize $f(\cdot)$ around the current mean of $q(\boldsymbol{x})$ using a first-order Taylor expansion, yielding a Gaussian approximation to the likelihood, but this can be inaccurate under strong curvature and requires Jacobians.

III. Sigma-Point Expectation Propagation

We now describe the proposed Sigma-Point EP (SP-EP) algorithm, which replaces first-order Taylor linearization inside EP by sigma-point integration. For clarity, we present SP-EP in a mean-field setting where the global approximation factorizes as

$$q(\boldsymbol{x}) = \prod_{i=1}^{n} q_i(x_i), \quad q_i(x_i) = \mathcal{N}(m_i, v_i). \tag{9}$$

This yields a diagonal covariance structure and keeps the per-iteration cost linear in n, while still allowing the likelihood to couple the components through the nonlinear map $f(\cdot)$.

We focus on updating the prior site associated with a single component x_i. The core idea is to:

- integrate out $\boldsymbol{x}_{-i}$ under the current Gaussian approximation using sigma points to obtain an extrinsic likelihood term $L_i(x_i)$, and
- update the Gaussian approximation to x_i by matching one-dimensional moments under the true prior $p_i(x_i)$.

A. Sigma-Point Extrinsic over $\boldsymbol{x}_{-i}$

Let $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$, and denote by

$$\boldsymbol{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)^\top, \tag{10}$$

and

$$q_{-i}(\boldsymbol{x}_{-i}) = \prod_{j\neq i} q_j(x_j) \tag{11}$$

the current Gaussian approximation over all variables except x_i. In the mean-field setting, $q_{-i}(\boldsymbol{x}_{-i})$ is Gaussian with mean $\boldsymbol{m}_{-i}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{-i,-i} = \operatorname{diag}(\boldsymbol{v}_{-i})$.

We define the *extrinsic* factor for x_i as

$$L_i(x_i) \triangleq \int p(\boldsymbol{y} \mid x_i, \boldsymbol{x}_{-i})\, q_{-i}(\boldsymbol{x}_{-i})\, \mathrm{d}\boldsymbol{x}_{-i}. \tag{12}$$

For the additive Gaussian noise model $\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{v}$ with $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$, this becomes

$$p(\boldsymbol{y} \mid x_i, \boldsymbol{x}_{-i}) = \mathcal{N}\big(\boldsymbol{y}; f(x_i, \boldsymbol{x}_{-i}), \boldsymbol{R}\big). \tag{13}$$

The integral in (12) is in general intractable. We approximate it using a sigma-point rule over $\boldsymbol{x}_{-i}$. Specifically, we instantiate the Unscented Transform (UT). Let $\{\boldsymbol{\Xi}_k^{(-i)}, W_k\}_{k=0}^{2(n-1)}$ denote the UT sigma points and weights for the Gaussian $q_{-i}(\boldsymbol{x}_{-i}) = \mathcal{N}(\boldsymbol{m}_{-i}, \boldsymbol{\Sigma}_{-i,-i})$, constructed with standard UT parameters (α, β, κ). For any fixed x_i, we approximate $L_i(x_i)$ by

$$L_i(x_i) \approx \sum_{k=0}^{2(n-1)} W_k\, \mathcal{N}\big(\boldsymbol{y}; f(x_i, \boldsymbol{\Xi}_k^{(-i)}), \boldsymbol{R}\big), \tag{14}$$

which yields a scalar function of x_i only.

B. Exact One-Dimensional Moments under the True Prior

Given the extrinsic $L_i(x_i)$, the exact (1D) tilted distribution for x_i under the true prior is

$$\hat{p}_i(x_i) \propto L_i(x_i)\, p_i(x_i). \tag{15}$$

We define the corresponding normalization constant and moments as

$$Z_i^\star = \int L_i(x_i) p_i(x_i)\, \mathrm{d}x_i, \tag{16}$$

$$\mu_i^\star = \frac{1}{Z_i^\star} \int x_i L_i(x_i) p_i(x_i)\, \mathrm{d}x_i, \tag{17}$$

$$(\sigma_i^2)^\star = \frac{1}{Z_i^\star} \int (x_i - \mu_i^\star)^2 L_i(x_i) p_i(x_i)\, \mathrm{d}x_i. \tag{18}$$

These one-dimensional integrals can be evaluated by Gauss–Hermite or Gauss–Legendre quadrature on a suitably chosen grid, or by light-weight Monte Carlo with cached sample points. In practice, we precompute a 1D grid for x_i (e.g., centered at the current m_i with span covering several standard deviations) and reuse it across iterations, caching the values of $L_i(x_i)$ where possible.

C. Gaussian Approximated Prior and Moment Matching

To maintain a Gaussian EP approximation, we introduce a Gaussian surrogate prior for x_i of the form

$$q_i(x_i) = \mathcal{N}(m_i, v_i), \tag{19}$$

with parameters (m_i, v_i) to be optimized. For any given (m_i, v_i), we can define the corresponding approximate tilted distribution

$$\tilde{p}_i(x_i; m_i, v_i) \propto L_i(x_i) q_i(x_i), \tag{20}$$

with normalization constant and moments

$$Z_i(m_i, v_i) = \int L_i(x_i) q_i(x_i)\, \mathrm{d}x_i, \tag{21}$$

$$\mu_i(m_i, v_i) = \frac{1}{Z_i(m_i, v_i)} \int x_i L_i(x_i) q_i(x_i)\, \mathrm{d}x_i, \tag{22}$$

$$\sigma_i^2(m_i, v_i) = \frac{1}{Z_i(m_i, v_i)} \int (x_i - \mu_i(m_i, v_i))^2 L_i(x_i) q_i(x_i)\, \mathrm{d}x_i. \tag{23}$$

These integrals are again one-dimensional and can be evaluated using the same quadrature grid used for the true prior moments, replacing $p_i(x_i)$ by $q_i(x_i)$.

We then determine (m_i, v_i) by matching the first two moments:

$$\mu_i(m_i, v_i) = \mu_i^\star, \tag{24}$$

$$\sigma_i^2(m_i, v_i) = (\sigma_i^2)^\star. \tag{25}$$

This yields a two-dimensional root-finding problem

$$F(m_i, v_i) = \begin{bmatrix} \mu_i(m_i, v_i) - \mu_i^\star \\ \sigma_i^2(m_i, v_i) - (\sigma_i^2)^\star \end{bmatrix} = 0, \tag{26}$$

which we solve using a damped Newton method. The required Jacobian entries, such as $\partial \mu_i / \partial m_i$ and $\partial \sigma_i^2 / \partial v_i$, can be obtained from derivatives of Gaussian weights in the quadrature or approximated by finite differences on the same grid. We initialize (m_i, v_i) at the current EP parameters and apply damping to ensure stability.

Algorithm 1 Sigma-Point EP (SP-EP) for Nonlinear, Non-Gaussian Inference

Input: Nonlinear map $f(\cdot)$, noise covariance $\boldsymbol{R}$, priors $\{p_i\}$, observation $\boldsymbol{y}$.
Initialize: Means $m_i^{(0)} = 0$, variances $v_i^{(0)} = v_0$, damping factor ρ, max iterations T.
for $t = 0$ to $T-1$ **do**
  for $i = 1$ to n **do**
    Form $q_{-i}(\boldsymbol{x}_{-i}) = \prod_{j \neq i} \mathcal{N}(x_j; m_j^{(t)}, v_j^{(t)})$.
    Construct UT sigma points $\{\boldsymbol{\Xi}_k^{(-i)}, W_k\}$ for q_{-i}.
    Build extrinsic $L_i(x_i)$ via (14) (cache L_i on a 1D grid).
    Compute $(Z_i^\star, \mu_i^\star, (\sigma_i^2)^\star)$ under true prior $p_i(x_i)$ by 1D quadrature.
    Solve for $(\check{m}_i, \check{v}_i)$ such that $\mu_i(\check{m}_i, \check{v}_i) = \mu_i^\star$ and $\sigma_i^2(\check{m}_i, \check{v}_i) = (\sigma_i^2)^\star$ using damped Newton, reusing the same grid and L_i.
    Update $m_i^{(t+1)} = (1-\rho) m_i^{(t)} + \rho \check{m}_i$, $v_i^{(t+1)} = (1-\rho) v_i^{(t)} + \rho \check{v}_i$, with variance floor $v_i^{(t+1)} \leftarrow \max(v_i^{(t+1)}, \epsilon)$ to maintain symmetric positive semidefinite (SPD).
  end for
  Check convergence of $\{m_i^{(t)}, v_i^{(t)}\}$; if converged, break.
end for
Output: Approximate posterior $q(\boldsymbol{x}) = \prod_i \mathcal{N}(x_i; m_i^{(T)}, v_i^{(T)})$.

D. EP Update and Damping

Let $(\check{m}_i, \check{v}_i)$ denote the solution of the moment-matching step for component x_i. We update the Gaussian approximate marginal for x_i via a damped step:

$$m_i \leftarrow (1-\rho) m_i + \rho \check{m}_i, \tag{27}$$

$$v_i \leftarrow (1-\rho) v_i + \rho \check{v}_i, \tag{28}$$

where $\rho \in (0, 1]$ is a damping factor. In our experiments, we use ρ between 0.1 and 0.5 depending on the SNR and curvature of $f(\cdot)$.

Algorithm 1 summarizes the proposed SP-EP procedure. We sweep over $i = 1, \ldots, n$ in a coordinate-wise fashion and iterate until changes in (m_i, v_i) are below prescribed thresholds or a maximum number of outer iterations is reached.

E. Complexity and Practical Considerations

We briefly analyze the computational cost of SP-EP per outer sweep. For each coordinate i we require:

- **Sigma-point propagation over $\boldsymbol{x}_{-i}$:** the UT over the $(n-1)$-dimensional cavity $q_{-i}(\boldsymbol{x}_{-i})$ uses $K = 2(n-1)+1$ sigma points $\{\boldsymbol{\Xi}_k^{(-i)}\}_{k=0}^{K-1}$. To build the extrinsic $L_i(x_i)$ on a one-dimensional grid $\{x_i^{(q)}\}_{q=1}^{Q}$, we evaluate the forward map $f(\cdot)$ for all combinations $(x_i^{(q)}, \boldsymbol{\Xi}_k^{(-i)})$, leading to QK calls to $f(\cdot)$ per coordinate i.

- **One-dimensional integration over x_i:** once $L_i(x_i^{(q)})$ has been cached on the grid, computing $Z_i^\star$, $\mu_i^\star$, and $(\sigma_i^2)^\star$ under the true prior $p_i(x_i)$, as well as $Z_i(m_i, v_i)$, $\mu_i(m_i, v_i)$, and $\sigma_i^2(m_i, v_i)$ under the Gaussian surrogate, only involves $\mathcal{O}(Q)$ scalar operations per Newton iteration. These are typically negligible compared to the cost of evaluating $f(\cdot)$.

Summing over all n coordinates, the dominant cost per outer sweep is on the order of

$$nQ[2(n-1)+1]$$

forward evaluations of the nonlinear map $f(\cdot)$, i.e., $\mathcal{O}(n^2 Q)$ in the state dimension n (for fixed grid size Q). Crucially, this complexity is Jacobian-free: the method treats $f(\cdot)$ as a black box. The remaining one-dimensional quadrature and Newton updates scale as $\mathcal{O}(nQ)$ and are inexpensive in the low-dimensional settings of interest.

To improve numerical stability, we adopt the following practical tricks:

- *Damping:* The Newton updates for (m_i, v_i) are damped, and the EP site updates are additionally relaxed with parameter ρ.
- *Variance floor and jitter:* We enforce a minimal variance $v_i \geq \epsilon$ for all i and add small jitter to $\boldsymbol{R}$ and to intermediate covariance estimates when needed to maintain positive definiteness.
- *Cavity scaling:* In challenging regimes (very low SNR or strong curvature), the cavity variance can become too small. We optionally inflate the cavity variances before constructing UT sigma points (e.g., by a scalar factor $c > 1$) and compensate in the site update, which improves robustness.
- *Cached grids:* One-dimensional grids and quadrature weights for x_i are precomputed and reused across iterations and even across components when the priors share parameters, reducing overhead.

The same interface readily accommodates alternative sigma-point rules such as CKF: one simply replaces the UT construction in (14) while keeping the EP shell and the one-dimensional moment-matching machinery unchanged.

IV. Numerical Experiments

We now illustrate the performance of SP-EP on a synthetic nonlinear model with non-Gaussian priors. We compare against a Taylor-based EP baseline and use a Metropolis–Hastings (MH) sampler as a reference to approximate the MMSE estimator.

A. Simulation Setup

We consider a three-dimensional state and measurement:

$$\boldsymbol{x} \in \mathbb{R}^3, \quad \boldsymbol{y} \in \mathbb{R}^3, \tag{29}$$

with nonlinear map

$$f(\boldsymbol{x}) = \begin{bmatrix} \sin(x_1) + 0.3x_2^2 \\ x_2 + 0.6x_1x_3 \\ 0.4x_1 + \tanh(x_3) \end{bmatrix}. \tag{30}$$

The measurement model is

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{v}, \quad \boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \sigma_v^2 \boldsymbol{I}_3). \tag{31}$$

We sweep the SNR over the set

$$\text{SNR} \in \{0, 4, 8, 12, 16, 20\} \text{ dB}, \tag{32}$$

by adjusting σ_v^2 accordingly.

We consider two classes of separable heavy-tailed priors namely Laplace and Student-*t* distributions. A Laplace prior with scale $b > 0$ reads

$$p_i(x_i) = \frac{1}{2b} \exp\left(-\frac{|x_i|}{b}\right), \tag{33}$$

while a zero-mean Student-*t* prior with degrees of freedom ν and scale $s > 0$ reads

$$p_i(x_i) = \frac{\Gamma\big((\nu+1)/2\big)}{\Gamma(\nu/2)\sqrt{\nu\pi}\, s} \left(1 + \frac{x_i^2}{\nu s^2}\right)^{-(\nu+1)/2}. \tag{34}$$

Such priors are commonly used to promote sparsity and robustness. In the numerical experiments, we will consider i.i.d. components with a Laplace prior using $b = 0.5$ and a Student-*t* prior using $\nu = 5$ and $s = 0.5$.

For each SNR level and each prior choice, we generate $M = 2000$ independent trials. In each trial, we draw $\boldsymbol{x}$ from the prior, generate $\boldsymbol{y}$ according to the nonlinear model, and then run the inference algorithms to estimate the posterior mean.

B. Baseline Methods

We compare the following methods:

- **SP-EP (UT):** The proposed Sigma-Point EP using UT for the sigma-point construction. We set the UT parameters to $(\alpha, \beta, \kappa) = (10^{-3}, 2, 0)$ and use a modest 1D quadrature grid (e.g., $Q = 15$ points) for the site integrals. Damping is applied with ρ in the range $[0.1, 0.3]$. We run up to $T = 1000$ outer sweeps or until convergence of $\{m_i, v_i\}$.
- **Taylor-EP:** A classical EP baseline in which we approximate the likelihood moments by first-order Taylor expansion of $f(\cdot)$ around the current mean $\hat{\boldsymbol{x}}$:

$$f(\boldsymbol{x}) \approx f(\hat{\boldsymbol{x}}) + \boldsymbol{J}(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}), \tag{35}$$

  where $\boldsymbol{J}(\hat{\boldsymbol{x}})$ is the Jacobian of f evaluated at $\hat{\boldsymbol{x}}$. This yields an approximate Gaussian likelihood and allows EP updates analogous to those in linear models. We assume access to $\boldsymbol{J}(\cdot)$ for this baseline and apply the same damping and stopping criteria as in SP-EP.
- **MH-MMSE:** As a reference, we run a random-walk Metropolis–Hastings sampler [6] targeting the exact posterior $p(\boldsymbol{x} \mid \boldsymbol{y})$ and approximate the MMSE estimator $\mathbb{E}[\boldsymbol{x} \mid \boldsymbol{y}]$ by sample averaging. We use 20 000 iterations with 5 000 burn-in steps per trial, with proposal covariance tuned to achieve an acceptance rate around 25%.

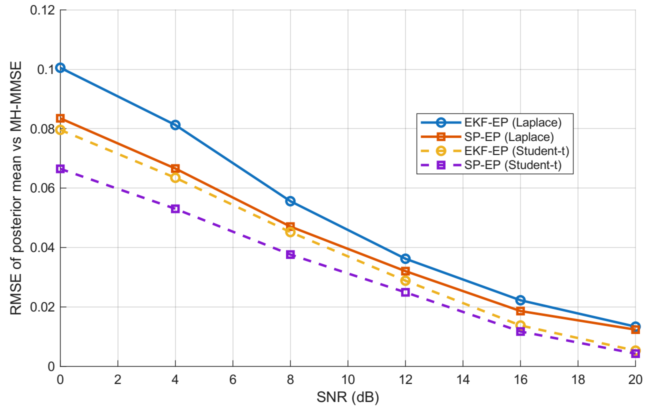

Fig. 1. RMSE of the posterior mean versus SNR for SP-EP (UT) and Taylor-EP under Laplace and Student-t priors. The reference is the MH-based MMSE estimator.

C. Performance Metric

For each method and each trial, we obtain an estimate $\hat{\boldsymbol{x}}$ of the posterior mean. Using the MH sampler as reference, we denote the approximate MMSE estimate by $\boldsymbol{x}^{\mathrm{MH}}$. We then compute the root-mean-square error (RMSE) of the posterior mean across trials:

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M} \|\hat{\boldsymbol{x}}^{(m)} - \boldsymbol{x}^{\mathrm{MH},(m)}\|_2^2}. \tag{36}$$

For each SNR and prior, we report the RMSE for SP-EP and Taylor-EP. To reduce the impact of outliers, we also check robust statistics such as median absolute deviation, although we omit those plots for brevity.

D. Results and Discussion

Fig. 1 shows the RMSE of the posterior mean versus SNR for SP-EP and Taylor-EP under Laplace and Student-t priors, with MH-MMSE as reference. Several observations can be made:

- Across all SNR levels and for both priors, SP-EP achieves lower RMSE than Taylor-EP.
- The performance gap is particularly pronounced at low SNR and under the heavy-tailed Student-t prior, where the nonlinearities and prior tails interact strongly. In these regimes, first-order Taylor EP tends to be biased and overly confident.
- As the SNR increases, both methods improve, and the gap narrows, but SP-EP still maintains a slight advantage due to its more accurate curvature-aware moment estimates.
- The computational overhead of SP-EP compared to Taylor-EP is moderate in this low-dimensional example. In higher dimensions, the cost of sigma-point integration grows with the number of sigma points, but remains attractive in settings where Jacobians are unavailable or expensive.

V. Conclusion

We have presented sigma-point expectation propagation (SP-EP), a Jacobian-free EP framework for nonlinear, non-Gaussian Bayesian inference. By embedding sigma-point extrinsics inside EP, SP-EP provides curvature-aware likelihood moments while treating the forward model as a black box, and retains the modular EP interface to heavy-tailed priors such as Laplace and Student-*t*. Numerical experiments on a three-dimensional nonlinear model show that SP-EP systematically improves posterior-mean RMSE over a Taylor-based EP baseline across a range of SNRs, while remaining computationally practical. Several extensions are of interest. One direction is to move from the static setting to sequential state-space models by embedding the sigma-point moment oracle into UKF-style filters and smoothers while retaining EP prior sites for non-Gaussianity. Another is to develop adaptive strategies for selecting sigma-point rules (e.g., UT versus CKF) based on local curvature and SNR. Finally, combining SP-EP with more expressive approximate-inference schemes, such as mixture-based message passing or variational methods, may help capture multimodal posteriors beyond the single-Gaussian regime considered here.

Acknowledgment

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, by the Franco-German project 5G-OPERA (BPI), the French project YACARI (PEPR-5G), the EU INFRA project CONVERGE, and by a Huawei France funded Chair towards Future Wireless Networks.

References

[1] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. Conf. Uncertainty in Artificial Intelligence (UAI)*, San Francisco, CA, USA, 2001, pp. 362–369.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[3] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[4] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation, and Controls*, Orlando, FL, USA, 1997, pp. 182–193.

[5] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[6] G. O. Roberts, A. Gelman, and W. R. Gilks, "Weak convergence and optimal scaling of random walk Metropolis algorithms," *Ann. Appl. Probab.*, vol. 7, no. 1, pp. 110–120, 1997.

[7] F. Xiao and D. Slock, "Breaking the Gaussian barrier: Leveraging ReG-VAMP to extend EKF and IEKF," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2024.

[8] M. Seeger, "Expectation propagation for exponential families," *Technical Report*, University of Edinburgh, 2005.

[9] J. M. Hernández-Lobato and R. P. Adams, "Probabilistic backpropagation for scalable learning of Bayesian neural networks," in *Proc. Int. Conf. Machine Learning (ICML)*, Lille, France, 2015, pp. 1861–1869.

[10] Z. Zhao, F. Xiao, and D. Slock, "Vector approximate message passing for not so large N.I.I.D. generalized I/O linear models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Seoul, South Korea, Apr. 2024.