State Estimation for Linear Systems with Non-Gaussian Measurement Noise via Dynamic Programming

Mohammad Hussein Yoosefian Nooshabadi and Laurent Lessard

***Abstract*— We propose a new recursive estimator for linear dynamical systems under Gaussian process noise and non-Gaussian measurement noise. Specifically, we develop an approximate *maximum a posteriori* (MAP) estimator using dynamic programming and tools from convex analysis. Our approach does not rely on restrictive noise assumptions and employs a Bellman-like update instead of a Bayesian update. Our proposed estimator is computationally efficient, with only modest overhead compared to a standard Kalman filter. Simulations demonstrate that our estimator achieves lower root mean squared error (RMSE) than the Kalman filter and has comparable performance to state-of-the-art estimators, while requiring significantly less computational power.**

I. Introduction

We consider state estimation for discrete-time linear systems driven by non-Gaussian noise, in the standard form

$$x_t = Ax_{t-1} + w_t, \tag{1a}$$
$$y_t = Cx_t + v_t, \tag{1b}$$

where A and C are known, and the distributions of the process noise w_t, measurement noise v_t, and initial state x_0 are specified and not necessarily Gaussian. We make the standard assumptions that w_t and v_t are independent of each other, of x_t, and across time.

In this setting, the exact posterior distribution given past measurements is provided by *Bayesian filtering*, which is a recursive formula for updating the conditional state distribution when a new measurement y_t arrives

$$\mathbb{P}(x_t \mid y_{0:t}) \propto \mathbb{P}(y_t \mid x_t) \int \mathbb{P}(x_t \mid x_{t-1})\, \mathbb{P}(x_{t-1} \mid y_{0:t-1})\, \mathrm{d}x_{t-1}. \tag{2}$$

The estimator that minimizes the mean squared error (MSE), is given by the conditional expectation

$$\hat{x}_t = \mathbb{E}(x_t \mid y_{0:t}) = \int x_t\, \mathbb{P}(x_t \mid y_{0:t})\, \mathrm{d}x_t. \tag{3}$$

When the noise and the initial state are Gaussian, the integral in (2) has a closed-form solution, and the estimator (3) can be found recursively via the celebrated Kalman filter (KF).

Despite the widespread applications of the KF [1]–[3], its Gaussianity assumption does not hold in many practical scenarios. Notable examples include underwater communication [4], power systems [5] and magnetic resonance imaging (MRI) [6], where Gaussian mixture or various heavy-tailed distributions have been reported. In the non-Gaussian setting, the integral in (2) typically does not have a closed-form solution, so one must resort to approximation techniques or alternatively, use estimators that do not directly deal with the integration (2). This leads to two general classes of estimators, which we briefly survey below.

Authors are with the Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115, USA. Email: {yoosefiannooshabad.m, l.lessard}@northeastern.edu

A. Integration-based Estimators

Directly approximating the integral in (2) yields a trade-off between computational effort and approximation error. At one extreme, we have *sequential Monte Carlo* approaches, such as bootstrap filtering [7] and particle filtering [8], where distributions are represented by a set of random samples. These approaches are computationally intensive and also applicable when the dynamics are nonlinear. At the other extreme, the simplest approach, which we call the *standard KF*, is to replace non-Gaussian noise distributions with Gaussian distributions of matching mean and variance and then apply a standard KF. Despite being computationally attractive, this approach may lead to severely degraded performance when the noise is heavy-tailed or bimodal.

Many other schemes have been proposed that balance computational load and approximation error. In [9] the noise and posterior distributions are approximated as a sum of Gaussians, yielding the *Gaussian sum filter*. Despite favorable performance compared to the standard KF, the number of Gaussians required to represent the posterior grows exponentially with the time horizon. The work [10] proposed a more scalable version of GSF. An alternative approach, proposed by Masreliez [11], approximates the Bayesian update using a *score function*; the gradient of $\log \mathbb{P}(y_t \mid y_{0:t-1})$. This method outperforms the standard KF, but requires evaluating a difficult integral similar to (2) at each timestep. The work [12] proposed an efficient implementation of the Masreliez estimator via polynomial approximation.

B. Optimization-based Estimators

Alternatively, one may forego the Bayesian update (2), and instead formulate an optimization problem that is solved whenever a new measurement arrives. For such estimators, conditional distributions are never computed or approximated; rather, robustness to non-Gaussianity is enforced by explicitly minimizing an auxiliary cost function. In [13], a *maximum correntropy Kalman filter* (MCKF) is proposed, which has shown superior performance compared to the standard KF in heavy-tailed noise environments. In [14] the *entropy* of the estimation error is minimized instead, which can outperform the MCKF in certain noise environments.

The authors in [15] develop optimization-based estimators that are resilient to bounded process noise and impulsive measurement noise. Finally, [16] minimizes the Kullback–Liebler divergence of the estimation error. Although these methods, when tuned properly, offer great performance, they are usually designed for specific measurement noise distributions. Furthermore, since these methods involve auxiliary optimization at each timestep, they tend to be computationally more demanding than the standard KF.

In this paper, we propose a new estimator inspired by *maximum a posteriori* (MAP) estimation. As we detail in Section III, the MAP framework yields a Bellman-style recursion rather than Bayesian recursion (2). Although this Bellman recursion does not have a closed-form solution, under proper approximation of the value function, it reduces to algebraic equations, which we use to develop a nonlinear estimator for the case with Gaussian process noise and non-Gaussian measurement noise. Simulations show that our proposed estimator achieves lower RMSE than the Kalman filter and matches the performance of state-of-the-art alternatives at a fraction of their computational cost.

The rest of the paper is organized as follows. We present preliminaries and assumptions in Section II, our main results in Section III, discussion in Section IV, and numerical examples in Section V. Finally, we conclude and discuss future research directions in Section VI.

II. Preliminaries and Assumptions

A. Notation

The notation $\mathbb{P}$ and $\mathbb{E}$ denote probability density function (PDF) and expectation of a random variable or vector. For a square symmetric real matrix M, we write $M \succ 0$ to mean that M is positive definite. We write $I_n \in \mathbb{R}^{n\times n}$ to denote the identity matrix. We write $\mathcal{N}(\mu, \Sigma)$ to denote a (multivariate) Gaussian distribution with mean μ and covariance $\Sigma \succ 0$. Also, $\ell(\cdot) := \log \mathbb{P}(\cdot)$ and $\ell(\,\cdot \mid \cdot\,) := \log \mathbb{P}(\,\cdot \mid \cdot\,)$ denote marginal and conditional log-likelihood, respectively.

For the rest of the paper, we use "$\star$" to denote submatrices that can be inferred from symmetry and "…" to denote entries omitted because they are irrelevant to the discussion.

B. Assumptions

We consider the dynamical system (1) where $x_t \in \mathbb{R}^n$, $y_t \in \mathbb{R}^d$, $w_t \in \mathbb{R}^n$, and $v_t \in \mathbb{R}^d$ are the state vector, measured output, process noise, and measurement noise, respectively, and A and C are known. For convenience, we reparameterize $\mathbb{P}(w_t) = e^{-q(w_t)}$ and $\mathbb{P}(v_t) = e^{-r(v_t)}$, where $q : \mathbb{R}^n \to \mathbb{R}$ and $r : \mathbb{R}^d \to \mathbb{R}$ are continuously differentiable. We also make the mild technical assumption that the mode exists and is unique for all distributions. These assumptions encompass many common distributions, such as Gaussian mixtures, Cauchy, and skewed normal.

C. The Kalman Filter

Consider the dynamical system (1). If the initial state and the noise inputs are Gaussian, $x_0 \sim \mathcal{N}(\mu_{0|-1}, P_{0|-1})$, $w_t \sim \mathcal{N}(\mu_w, \Sigma_w)$ and $v_t \sim \mathcal{N}(\mu_v, \Sigma_v)$, then the state has a Gaussian posterior $(x_t \mid y_0, \ldots, y_t) \sim \mathcal{N}(\mu_t, P_t)$, which can be computed recursively via the celebrated Kalman filter (KF). The two-step KF can be written as follows [17, §6.2].

- *Time update:*

$$P_{t|t-1} = \Sigma_w + A P_{t-1} A^\top \tag{4a}$$

$$\mu_{t|t-1} = A\mu_{t-1} + \mu_w \tag{4b}$$

- *Measurement update:*

$$P_t^{-1} = P_{t|t-1}^{-1} + C^\top \Sigma_v^{-1} C \tag{5a}$$

$$\mu_t = \mu_{t|t-1} + P_t C^\top \Sigma_v^{-1}(y_t - C\mu_{t|t-1} - \mu_v) \tag{5b}$$

D. The Fenchel Conjugate

The Fenchel conjugate of a function, also known as the convex conjugate or simply conjugate, is defined as follows.

Definition 1: For a function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the conjugate function g^* is defined as [18, §3.3]

$$g^*(\lambda) = \sup_x \big(\lambda^\top x - g(x)\big).$$

Some useful properties: the conjugate g^* is a always a convex function, and if g itself is convex, then $g^{**} := (g^*)^* = g$.

Proposition 1: Given a matrix $M \succ 0$, vector m and scalar γ of compatible dimensions, the conjugate of a convex quadratic function is given as follows.

If $g(x) = \dfrac{1}{2}\begin{bmatrix} x \\ 1 \end{bmatrix}^\top \begin{bmatrix} M & m \\ m^\top & \gamma \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$, then

$$g^*(\lambda) = \frac{1}{2}\begin{bmatrix} \lambda \\ 1 \end{bmatrix}^\top \begin{bmatrix} M^{-1} & -M^{-1}m \\ -m^\top M^{-1} & m^\top M^{-1} m - \gamma \end{bmatrix} \begin{bmatrix} \lambda \\ 1 \end{bmatrix}.$$

E. Dynamic Programming for MAP Estimation

It was recently observed by Lange [19] that MAP estimators satisfy a Bellman-style recursion different from the Bayesian recursion (2). We briefly review this result. Consider the dynamical system (1). The MAP estimator for x_t given measurements $y_0, \ldots, y_t$ is found by maximizing the log-posterior $\log \mathbb{P}(x_{0:t} \mid y_{0:t})$, which is equivalent to maximizing the joint log-likelihood of the states and measurements, denoted $L(x_{0:t}, y_{0:t})$.[1] By the probability chain rule

$$L(x_{0:t}, y_{0:t}) = \ell(x_0) + \sum_{i=1}^{t} \ell(x_i \mid x_{i-1}) + \sum_{i=0}^{t} \ell(y_i \mid x_i). \tag{6}$$

Definition 2: The *value function* $V_t : \mathbb{R}^n \to \mathbb{R}$ is the joint negative log-likelihood minimized over past states

$$V_t(x) = \min_{x_0, \ldots, x_{t-1}} -L(x_{0:t-1}, x, y_{0:t}). \tag{7}$$

Using (6) and (7) we can obtain a Bellman-style recursion.

Lemma 1 (Bellman recursion): The value function given in (7) satisfies the following forward recursion

$$V_t(x) = -\ell(y_t \mid x) + \min_\xi \big\{ -\ell(x \mid \xi) + V_{t-1}(\xi) \big\}. \tag{8}$$

[1] This is the *trajectory* or *batch* MAP, which maximizes the conditional distribution of the full state history. Some authors also consider the *pointwise* or *filtering* MAP, which instead maximizes the conditional distribution of the current state. One way to compute this pointwise MAP estimate is to use the Bayesian update (2) but replace the expectation in (3) with an argmax.

In the Gaussian setting (see Section II-C) the value function is the convex quadratic

$$V_t(x) = \frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top \begin{bmatrix}P_t^{-1} & -P_t^{-1}\mu_t\\ \star & \dots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}, \tag{9}$$

where μ_t and $P_t \succ 0$ are the mean and the covariance of the posterior at time t. In this setting, the Bellman update (8), just like the Bayesian update (2), reduces to the standard KF.

III. Main Results

Our starting point is the Bellman recursion of Lemma 1.

A. Quadratic approximation

In this paper, we approximate the value function (7) with a quadratic function of the form (9). The justification for this approximation is that many log-PDFs resemble quadratic functions near their modes [19]. Quadratic value functions also reduce Bellman recursions to a pair of algebraic equations via the Fenchel dual, a fact observed in [20].

Lemma 2: Consider the linear system (1) with $q(\cdot)$ and $r(\cdot)$ defined in Section II-B. If V_t and V_{t-1} are quadratic functions of the form (9), then there exists a function $p(\cdot)$ such that the Bellman recursion (8) simplifies to the following pair of algebraic equations

$$p(x) + r(y_t - Cx) = \frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top \begin{bmatrix}P_t^{-1} & -P_t^{-1}\mu_t\\ \star & \dots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}, \tag{10a}$$

$$p^*(\lambda) - q^*(\lambda) = \frac{1}{2}\begin{bmatrix}\lambda\\1\end{bmatrix}^\top \begin{bmatrix}AP_{t-1}A^\top & A\mu_{t-1}\\ \star & \dots\end{bmatrix}\begin{bmatrix}\lambda\\1\end{bmatrix}. \tag{10b}$$

Proof: See Section A. ■

Corollary 1: In the setting of Lemma 2, if we further assume the process noise is Gaussian with $w_t \sim \mathcal{N}(\mu_w, \Sigma_w)$, then Eq. (10) simplifies to the single equation

$$\begin{aligned}&\frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top \begin{bmatrix}P_t^{-1} & -P_t^{-1}\mu_t\\ \star & \dots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}\\ &= \frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top \begin{bmatrix}P_{t|t-1}^{-1} & -P_{t|t-1}^{-1}\mu_{t|t-1}\\ \star & \dots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix} + r(y_t - Cx),\end{aligned} \tag{11}$$

where $P_{t|t-1}$ and $\mu_{t|t-1}$ are given in (4).

Proof: See Section B. ■

B. Proposed Estimator

For the rest of this paper, we assume the process noise is Gaussian, i.e., q is quadratic, and use Corollary 1 to derive our estimator. The more general case of non-Gaussian process noise is left for future studies.

Our proposed estimator is based on using the quadratic approximation for the value function described in Section III-A, together with a specialized quadratic approximation of r.

We propose a quadratic approximation for r *about a given point* $\bar{v}$ in the following sense. We choose the Hessian M_r of the approximation by fitting a quadratic whose gradient matches that of r at $\bar{v}$ and whose vertex is at the *minimizer of* r (the mode m_v of the measurement noise distribution). This leads to the equation $M_r(\bar{v} - m_v) = \nabla r(\bar{v})$. Among possible choices of M_r, we opt for the simple diagonal solution

$$M_r = \operatorname{diag}\left(\frac{[\nabla r(\bar{v})]_i}{\bar{v}_i - [m_v]_i}\right), \tag{12}$$

where $[x]_i$ denotes the i^{th} component of the vector x.

We require $M_r \succ 0$, which will typically be true, but may not be the case if the distribution is bimodal or has finite support. For such cases, we make some sensible adjustments, which we detail in Remarks 4 and 5. The complete quadratic approximation therefore has the form

$$r(v) \approx \frac{1}{2}(v - m_v)^\top M_r (v - m_v) + \text{(constant)}, \tag{13}$$

with M_r defined in (12). To approximate $r(v)$ in (11) we use $\bar{v} := y_t - C\mu_{t|t-1}$, yielding the approximation

$$r(y_t - Cx) \approx \frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top \begin{bmatrix}C^\top M_r C & -C^\top M_r(y_t - m_v)\\ \star & \dots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}.$$

So, rather than approximating $r(v)$ globally as in the standard KF, we use an approximation that is more accurate *near our current estimate of* v. Substituting the approximation of $r(y_t - Cx)$ into (11) and simplifying yields our proposed estimator, given below.

Proposed estimator. Given prior parameters μ_{t-1}, P_{t-1} and new measurement y_t, we find μ_t, P_t via the updates

$$P_{t|t-1} = \Sigma_w + AP_{t-1}A^\top \tag{14a}$$

$$\mu_{t|t-1} = A\mu_{t-1} + \mu_w \tag{14b}$$

$$P_t^{-1} = P_{t|t-1}^{-1} + C^\top M_r C, \tag{14c}$$

$$\mu_t = \mu_{t|t-1} + P_t C^\top \nabla r(\bar{v}), \tag{14d}$$

with $\bar{v} = y_t - C\mu_{t|t-1}$ and M_r given in (12).

Remark 1: Since we assume Gaussian process noise, Eqs. (14a) and (14b) are the same as the time update (4) of the standard KF. When the measurement noise is Gaussian with a diagonal covariance matrix, Eqs. (14c) and (14d) reduce to the measurement update (5) of the standard KF.

Remark 2: Consider the optimization problem

$$\min_\xi \left\{\frac{1}{2}(\xi - \mu_{t|t-1})^\top P_{t|t-1}^{-1}(\xi - \mu_{t|t-1}) + r(y_t - C\xi)\right\}, \tag{15}$$

where the objective is the negative log-posterior distribution $\ell(x_t = \xi \mid y_{0:t}) = \ell(x_t = \xi \mid y_{0:t-1}) + \ell(y_t \mid x_t = \xi)$. For Gaussian process noise, the prior $\mathbb{P}(x_t = \xi \mid y_{0:t-1})$ is Gaussian. Using our quadratic approximation for $r(y_t - C\xi)$ about the point $\xi = \mu_{t|t-1}$ and substituting into (15), and solving for ξ yields a *Newton-like* update that precisely recovers our estimator (14). One can envision other possible estimators that perform multiple Newton-like steps per measurement, or use different iterative schemes altogether.

Remark 3: Since our proposed approach is similar to a Newton step (see Remark 2), one may be tempted to use the second-order Taylor approximation $M_r = \nabla^2 r(\bar{v})$ and obtain an exact Newton step. However, Newton's method is prone to numerical issues when the Hessian is indefinite or

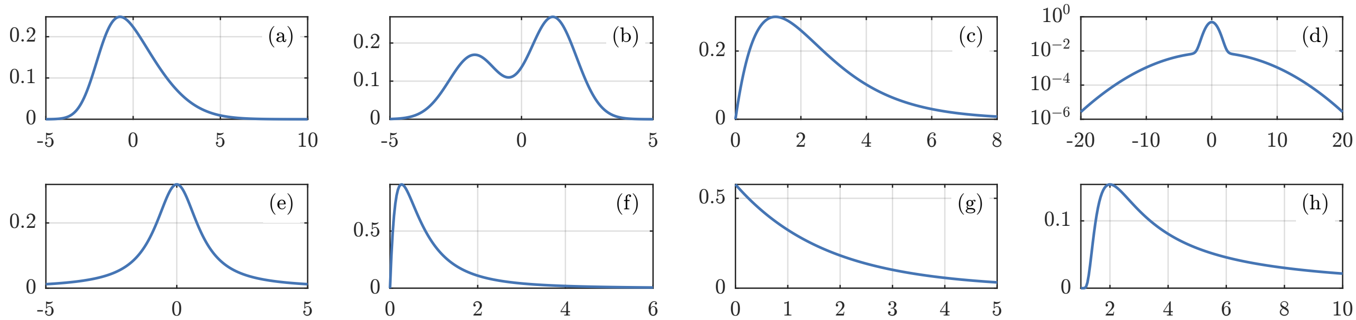

Fig. 1: PDFs of measurement noise distributions used in our numerical experiments: **(a)** Skewed normal, **(b)** Bimodal Gaussian mixture, **(c)** Gamma, **(d)** Impulsive Gaussian mixture (on a log scale), **(e)** Cauchy, **(f)** Beta prime, **(g)** Exponential, and **(h)** Lévy distributions. Refer to Table I for exact PDFs.

nearly singular (as in bimodal or heavy-tailed distributions). Anchoring the quadratic approximation to the mode of the distribution as in (13) gives our estimator added robustness.

IV. Discussion

Our estimator (14) has similar update equations to those of the Masreliez estimator [11], which uses a score function. However, the Masreliez estimator is fundamentally different because it approximates the Bayesian update (2), while our approach uses the Bellman update (8). The Masreliez estimator directly approximates the posterior distribution as a Gaussian, while we approximate the *noise distribution* at each iteration based on the most recent state estimate. As a result, our estimator is comparable to the standard KF in terms of computational footprint, while the Masreliez estimator is significantly more expensive. Specifically, in the Masreliez filter, in order to compute the score function at each timestep, the density of the predicted measurement needs to be computed via convolution, i.e.,

$$\mathbb{P}(y_t \mid y_{0:t-1}) = \int \mathbb{P}(y_t \mid x_t)\, \mathbb{P}(x_t \mid y_{0:t-1})\, \mathrm{d}x_t, \qquad (16)$$

where the prior $\mathbb{P}(x_t \mid y_{0:t-1})$ is assumed to be Gaussian. A common method to compute integrals of the form (16) is Gauss–Hermite quadrature, which has the computation complexity of $O(s^n)$, where s is the number of points used to grid the integration domain. As a result, the total complexity of the Masreliez estimator is $O(n^3 + s^n)$, while the complexity of our method is $O(n^3)$, which is the same as the standard KF. As we will see in Section V, our estimator has comparable performance to the Masreliez filter in terms of RMSE, while requiring significantly less computation.

For a linear dynamical system with non-Gaussian noise, the Kalman filter is the *best linear unbiased estimator* (BLUE) [17, §5.2]. Our proposed estimator (14) is different from a standard KF where noise is *globally* approximated by a Gaussian. Rather, we use *local* approximations of the noise distribution at each timestep, which leads to a nonlinear estimator and improved empirical performance compared to the standard KF (see Section V).

V. Numerical Examples

We evaluated the performance of our proposed estimator (14) on the following 2D rotational system [13]

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{18} & -\sin\frac{\pi}{18} \\ \sin\frac{\pi}{18} & \cos\frac{\pi}{18} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}, \qquad (17a)$$

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} x_t + v_t. \qquad (17b)$$

For the process noise, we used $w_t \sim \mathcal{N}(0, 0.05 I_2)$. For the measurement noise, we used $v_t \sim \mathcal{N}(0, 3)$ as our baseline. We then tested a variety of measurement noise distributions (see Fig. 1 and Table I). To ensure consistency across experiments, our noise PDFs were adjusted so that their first and second moments, matched those of the baseline measurement noise whenever possible.

For each noise distribution, we performed 200 Monte Carlo runs with different noise realizations, each time simulating (17) for 200 timesteps and reporting the mean and standard deviation of the final RMSE and the geometric mean of computation time. We compared our proposed estimator to the standard KF, the Masreliez filter [11] and the MCKF [13]. We also ran a particle filter (PF), which served as a minimum MSE baseline. For methods with tuning parameters, we tried different tunings and chose the most performant one. Specifically, for the PF we used 1000 particles and for the MCKF we used a kernel size of 5. For the Masreliez filter, the integral in (16) was evaluated numerically over a domain spanning ± 5 standard deviations from the prior estimate, discretized into a grid of 50 points. Estimators were implemented in MATLAB and run on a conventional laptop.

Simulation results are shown in Table II. We observe that our estimator outperforms the standard KF under all measurement noise distributions, while having comparable computation time. The Masreliez filter performs slightly better than our estimator, however its computation time is significantly higher than that of our proposed estimator.

With the exception of the standard KF, all the filters we tested were nonlinear and provided little in the way of theoretical guarantees, such as stability of the error dynamics. Indeed, the Masreliez filter was found to diverge in certain cases with heavy-tailed measurement noise.
A typical RMSE plot is shown in Fig. 2.

Remark 4: Special consideration is required when computing M_r for the bimodal distribution (b), as the Eq. (12)

TABLE I: Various noise distributions used in numerical experiments. All distributions are configured to share a common mean and variance whenever possible. In this table, $\phi(\cdot)$ and $\Phi(\cdot)$ represent the PDF and cumulative distribution function (CDF) of a standard Gaussian distribution, respectively. Furthermore, $\Gamma(\cdot)$ and $B(\cdot,\cdot)$ denote the gamma and beta functions, respectively. These PDFs are plotted in Fig. 1.

Noise type	Skew normal (a)	Bimodal (b)	Gamma (c)	Impulsive (d)
Noise PDF	$\frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\,\frac{x-\xi}{\omega}\right)$	$\sum_{i=1}^{2}\alpha_i\,\phi\left(\frac{x-\mu_i}{\sigma_i}\right)$	$\frac{1}{\Gamma(\alpha)\theta^\alpha}\,x^{\alpha-1}\exp\left(\frac{-x}{\theta}\right)$	$\sum_{i=1}^{2}\alpha_i\,\phi\left(\frac{x-\mu_i}{\sigma_i}\right)$
Parameters	$\xi=-2.0063$ $\omega=2.6505$ $\alpha=3$	$\alpha_1=0.4$ $\alpha_2=0.6$ $\mu_1=-1.8$ $\mu_2=1.2$ $\sigma_1^2=0.9$ $\sigma_2^2=0.8$	$\alpha=2$ $\theta=\sqrt{3/2}$	$\alpha_1=0.1$ $\alpha_2=0.9$ $\mu_1=0$ $\mu_2=0$ $\sigma_1^2=25$ $\sigma_2^2=0.5556$
Noise type	Cauchy (e)	Beta prime (f)	Exponential (g)	Lévy (h)
Noise PDF	$\frac{1}{\pi}\,\frac{\gamma}{(x-x_0)^2+\gamma^2}$	$\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}$	$\lambda\exp(-\lambda x)$	$\frac{\sqrt{c/(2\pi)}}{(x-\mu)^{3/2}}\exp\left(\frac{-c}{2(x-\mu)}\right)$
Parameters	$x_0=0$ $\gamma=1$	$\alpha=2$ $\beta=2.7891$	$\lambda=1/\sqrt{3}$	$\mu=1$ $c=3$

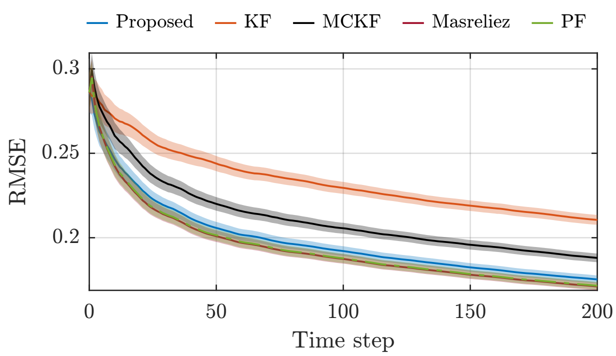

Fig. 2: Performance of the proposed estimator (14) compared with a variety of estimators under Gaussian process noise and impulsive measurement noise (Item (d) in Fig. 1 and Table I). We plot the mean RMSE of 1000 trials. The shaded area shows the 95% confidence interval for the mean.

assumes a single mode. To resolve this ambiguity, we use the direction of the gradient at $\bar{v}$. When $\bar{v}$ is between the two modes, for each i, we choose m_v to be the mode such that $[\nabla r(\bar{v})]_i/(\bar{v}_i-[m_v]_i)>0$, thus ensuring $M_r\succ 0$.

Remark 5: For distributions with compact support, whenever $\bar{v}$ falls outside of the valid domain, we perform a projection with ε-margin. For example, the support of the Lévy distribution is (μ,∞). Thus, we use $\bar{v}=\max(\bar{v},\mu+\varepsilon)$, where ε is a vector of small positive entries.

Code to generate all figures and tables can be found here: `https://github.com/QCGroup/dpkf`.

VI. Conclusion and Future Work

In this paper, we studied MAP estimation for linear systems under non-Gaussian noise from an optimization perspective. We showed that while the general Bellman recursion for MAP estimation does not have a closed-form solution, it can be reduced to algebraic equations via approximation of the value function. To locally approximate noise distributions, we introduced a novel robust quadratic approximation. These results were then leveraged to design a nonlinear state estimator under Gaussian process and non-Gaussian measurement noise. Simulations demonstrated that our proposed estimator outperformed the standard KF and had comparable performance to state-of-the-art estimators, highlighting its practical effectiveness. Moreover, our proposed estimator is computationally efficient, requiring only modest overhead compared to the standard Kalman Filter.

The estimator designed in this paper was for systems with Gaussian process and non-Gaussian measurement noise. One immediate extension is to use Lemma 2 to derive estimators for cases in which the process noise is also non-Gaussian. Another future direction is generalizing Lemma 2 to derive smoothers and predictors based on the Bellman recursion (8), or to generalize our work to nonlinear dynamical systems and develop analogs of the extended or unscented Kalman filter. Finally, more extensive numerical and theoretical analyses are needed to better characterize the performance and limitations of our proposed estimator.

Appendix

A. Proof of Lemma 2

Substituting the approximation (9) for V_t and V_{t-1} and $\ell(y_t\mid x)=-r(y_t-Cx)$ and $\ell(x\mid\xi)=-q(x-A\xi)$ into the Bellman recursion (8), we obtain

$$\frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top\begin{bmatrix}P_t^{-1} & -P_t^{-1}\mu_t\\ \star & \cdots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}=r(y_t-Cx)$$
$$+\min_\xi\left\{q(x-A\xi)+\frac{1}{2}\begin{bmatrix}\xi\\1\end{bmatrix}^\top\begin{bmatrix}P_{t-1}^{-1} & -P_{t-1}^{-1}\mu_{t-1}\\ \star & \cdots\end{bmatrix}\begin{bmatrix}\xi\\1\end{bmatrix}\right\}. \tag{18}$$

We then define

$$p(x):=\frac{1}{2}\begin{bmatrix}x\\1\end{bmatrix}^\top\begin{bmatrix}P_t^{-1} & -P_t^{-1}\mu_t\\ \star & \cdots\end{bmatrix}\begin{bmatrix}x\\1\end{bmatrix}-r(y_t-Cx),$$

which rearranges to (10a). Expressing (18) in terms of $p(x)$,

$$p(x)=\min_\xi\left\{q(x-A\xi)+\frac{1}{2}\begin{bmatrix}\xi\\1\end{bmatrix}^\top\begin{bmatrix}P_{t-1}^{-1} & -P_{t-1}^{-1}\mu_{t-1}\\ \star & \cdots\end{bmatrix}\begin{bmatrix}\xi\\1\end{bmatrix}\right\}.$$

Taking the conjugate of both sides via Definition 1 and simplifying, we obtain (10b), which completes the proof. ■

TABLE II: Simulation results for the system (17) with Gaussian process noise and various measurement noise distributions (see Fig. 1 and Table I). We report the mean and standard deviation of the RMSE over 200 Monte Carlo runs, and the geometric mean runtime, normalized so that KF = 1.00.

	Skewed normal (a)		Bimodal (b)		Gamma (c)		Impulsive (d)	
Estimator	RMSE ± s.d.	Time	RMSE ± s.d.	Time	RMSE ± s.d.	Time	RMSE ± s.d.	Time
KF	0.213 ± 0.050	1.00	0.217 ± 0.051	1.00	0.213 ± 0.045	1.00	0.212 ± 0.048	1.00
MCKF	0.212 ± 0.050	6.45	0.318 ± 0.064	6.63	0.212 ± 0.045	6.48	0.191 ± 0.043	8.20
Masreliez	0.206 ± 0.049	56.48	0.202 ± 0.045	39.87	diverged	—	0.173 ± 0.038	37.28
Ours (14)	0.205 ± 0.049	0.93	0.209 ± 0.048	1.47	0.198 ± 0.043	0.82	0.177 ± 0.038	1.34
PF	0.205 ± 0.049	63.67	0.201 ± 0.045	65.27	0.191 ± 0.041	76.02	0.173 ± 0.038	62.95

	Cauchy (e)		Beta prime (f)		Exponential (g)		Lévy (h)	
Estimator	RMSE ± s.d.	Time	RMSE ± s.d.	Time	RMSE ± s.d.	Time	RMSE ± s.d.	Time
KF	N/A	—	0.204 ± 0.048	1.00	0.212 ± 0.049	1.00	N/A	—
MCKF†	N/A	—	0.201 ± 0.047	6.82	0.206 ± 0.046	8.90	N/A	—
Masreliez	0.216 ± 0.047	28.49	diverged	—	diverged	—	diverged	—
Ours (14)	0.228 ± 0.052	1.00	0.184 ± 0.043	0.92	0.203 ± 0.050	0.90	0.254 ± 0.058	1.00
PF	0.216 ± 0.048	51.71	0.156 ± 0.034	80.25	0.177 ± 0.040	77.73	0.242 ± 0.053	76.88

† MCKF's derivation assumes finite noise covariance, which is undefined for Cauchy and Lévy distributions. In practice, one can still use it in these cases, but it requires manually tuning a nominal covariance parameter.

B. *Proof of Corollary 1*

Since $w \sim \mathcal{N}(\mu_w, \Sigma_w)$ we have

$$q(w) = \frac{1}{2}(w-\mu_w)^\top \Sigma_w^{-1}(w-\mu_w) + \text{(constant)}. \quad (19)$$

Taking the conjugate of (19) using Proposition 1 we get

$$q^*(\lambda) = \frac{1}{2}\begin{bmatrix}\lambda \\ 1\end{bmatrix}^\top \begin{bmatrix}\Sigma_w & \mu_w \\ \star & \dots\end{bmatrix}\begin{bmatrix}\lambda \\ 1\end{bmatrix}. \quad (20)$$

Substitute (20) into (10b) and use (4) to obtain

$$p^*(\lambda) = \frac{1}{2}\begin{bmatrix}\lambda \\ 1\end{bmatrix}^\top \begin{bmatrix}P_{t|t-1} & \mu_{t|t-1} \\ \star & \dots\end{bmatrix}\begin{bmatrix}\lambda \\ 1\end{bmatrix}. \quad (21)$$

Taking the conjugate of (21) using Proposition 1 yields

$$p(x) = \frac{1}{2}\begin{bmatrix}x \\ 1\end{bmatrix}^\top \begin{bmatrix}P_{t|t-1}^{-1} & -P_{t|t-1}^{-1}\mu_{t|t-1} \\ \star & \dots\end{bmatrix}\begin{bmatrix}x \\ 1\end{bmatrix}. \quad (22)$$

Substituting (22) into (10a) yields (11), as required. ■

References

[1] Q. Han, L. Yang, J. Du, and L. Cheng, "Blind equalization for chaotic signals based on echo state network and Kalman filter under nonlinear channels," *IEEE Communications Letters*, vol. 25, no. 2, pp. 589–592, 2020.

[2] M. Abdollahi, S. H. Pourtakdoust, M. Y. Nooshabadi, and H. N. Pishkenari, "An improved multi-state constraint Kalman filter for visual-inertial odometry," *Journal of the Franklin Institute*, vol. 361, no. 15, p. 107130, 2024.

[3] M. Taassori and B. Vizvári, "Enhancing medical image denoising: A hybrid approach incorporating adaptive Kalman filter and non-local means with Latin square optimization," *Electronics*, vol. 13, no. 13, p. 2640, 2024.

[4] J. Wang, J. Li, S. Yan, W. Shi, X. Yang, Y. Guo, and T. A. Gulliver, "A novel underwater acoustic signal denoising algorithm for Gaussian/non-Gaussian impulsive noise," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 429–445, 2020.

[5] V. Sarfi, A. Ghasemkhani, I. Niazazari, H. Livani, and L. Yang, "Decentralized dynamic state estimation with bimodal Gaussian mixture measurement noise," in *2019 North American Power Symposium (NAPS)*. IEEE, 2019, pp. 1–5.

[6] S. St-Jean, A. De Luca, M. A. Viergever, and A. Leemans, "Automatic, fast and robust characterization of noise distributions for diffusion MRI," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 304–312.

[7] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE proceedings F (radar and signal processing)*, vol. 140, no. 2. IET, 1993, pp. 107–113.

[8] J. Elfring, E. Torta, and R. Van De Molengraft, "Particle filters: A hands-on tutorial," *Sensors*, vol. 21, no. 2, p. 438, 2021.

[9] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.

[10] I. Bilik and J. Tabrikian, "MMSE-based filtering in presence of non-Gaussian system and measurement noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1153–1170, 2010.

[11] C. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 107–110, 1975.

[12] W.-R. Wu and A. Kunda, "Kalman filtering in non-Gaussian environment using efficient score function approximation," in *1989 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 1989, pp. 413–416.

[13] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.

[14] J. He, G. Wang, H. Yu, J. Liu, and B. Peng, "Generalized minimum error entropy Kalman filter for non-Gaussian noise," *ISA transactions*, vol. 136, pp. 663–675, 2023.

[15] A. Kircher, L. Bako, E. Blanco, and M. Benallouch, "An optimization framework for resilient batch estimation in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 10, pp. 5246–5261, 2021.

[16] S. Gultekin and J. Paisley, "Nonlinear Kalman filtering with divergence minimization," *IEEE Transactions on Signal Processing*, vol. 65, no. 23, pp. 6319–6331, 2017.

[17] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[18] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[19] R.-J. Lange, "Bellman filtering and smoothing for state–space models," *Journal of Econometrics*, vol. 238, no. 2, p. 105632, 2024.

[20] B. Hassibi, J. Hajar, and R. Ghane, "Beyond quadratic costs in LQR: Bregman divergence control," *arXiv preprint arXiv:2505.00317*, 2025.