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

Mohammad H. Yoosefian Nooshabadi
Northeastern University

Laurent Lessard
Northeastern University

IEEE Conference on Decision and Control

December 9–12, 2025

Estimation with non-Gaussian measurement noise

$$x_{t+1} = Ax_t + w_t$$
$$y_t = Cx_t + v_t$$

- A, C known; w_t, v_t i.i.d. with known pdfs
- Process noise w_t: Gaussian
- Measurement noise v_t: non-Gaussian.
  - Impulsive (underwater acoustic denoising)
  - Bimodal (power systems)
  - Heavy-tailed (MRI imaging)
  - Skewed (radar tracking)
  - Bounded support (wireless communications)

Estimation with non-Gaussian measurement noise

$$x_{t+1} = Ax_t + w_t$$
$$y_t = Cx_t + v_t$$

Goal: Design a recursive, online state estimator for non-Gaussian measurement noise that approximately minimizes the mean squared error (MMSE).

Kalman filter?

$$x_{t+1} = Ax_t + w_t$$
$$y_t = Cx_t + v_t$$

- Gaussian noise: KF is the optimal MMSE estimator.
- non-Gaussian noise: KF is the optimal linear MMSE estimator.

A nonlinear estimator can outperform KF

Bayesian filtering recursion

- Posterior update at time t (exact Bayesian recursion):

$$\underbrace{p(x_t \mid y_{0:t})}_{\text{new posterior}} \propto \underbrace{p(y_t \mid x_t)}_{\text{measurement}} \int \underbrace{p(x_t \mid x_{t-1})}_{\substack{\text{process}\\ \text{(Gaussian)}}} \underbrace{p(x_{t-1} \mid y_{0:t-1})}_{\text{old posterior}} \, \mathrm{d}x_{t-1}.$$

- MMSE estimator: $\hat{x}_t^{\mathsf{MMSE}} = \mathbb{E}[x_t \mid y_{0:t}]$.
- MAP estimator: $\hat{x}_t^{\mathsf{MAP}} = \arg\max_x p(x_t \mid y_{0:t})$.

- Linear $+$ Gaussian $\Rightarrow$ all terms Gaussian $\Rightarrow$ Kalman filter.
- Non-Gaussian v_t $\Rightarrow$ typically no closed-form expressions.

Popular approaches

Approximate the Bayesian recursion:

- Monte Carlo: Particle filter, bootstrap filter
- Gaussian sum filter
- Masreliez filter (score function approximation)
- Kalman filters (standard, H_∞, risk-sensitive)

expensive + asymptotically optimal

cheap + suboptimal

Solve auxiliary optimization:

- Maximum correntropy Kalman filter (MCKF)
- KL-divergence, generalized entropy,...

medium cost + distribution-specific

Key idea

Replace MMSE estimation with batch MAP estimation

- satisfies dynamic programming recursion
- easier to approximate than Bayesian recursion
- Gaussian case: recovers the KF
- non-Gaussian case: works pretty well!

Dynamic programming view of (batch) MAP

- Define value function[1]

$$V_t(x) \;=\; \min_{x_0,\ldots,x_{t-1}} \; -\log p(x_{0:t-1}, x, y_{0:t})$$

- Satisfies a (forward) Bellman recursion:

$$V_t(x) \;=\; \underbrace{-\log p(y_t \mid x)}_{\text{measurement}} + \min_{\xi} \Big\{ \underbrace{-\log p(x \mid \xi)}_{\substack{\text{process} \\ \text{(quadratic)}}} + V_{t-1}(\xi) \Big\}.$$

- Batch MAP estimate: $\hat{x}_t^{\mathsf{MAP}} = \arg\min_x V_t(x)$.

[1] R.-J. Lange, "Bellman filtering and smoothing for state-space models," Journal of Econometrics (2024).

Quadratic approximation idea

$$V_t(x) \;=\; \underbrace{-\log\, p(y_t \mid x)}_{\text{measurement}} + \min_{\xi} \Big\{ \underbrace{-\log\, p(x \mid \xi)}_{\substack{\text{process}\\ \text{(quadratic)}}} + V_{t-1}(\xi) \Big\}$$

- Approximate $V_t(x)$ as a quadratic (μ_t is the MAP estimate):

$$V_t(x) \approx \tfrac{1}{2}(x - \mu_t)^\mathsf{T} P_t^{-1}(x - \mu_t)$$

- Replace $-\log p(y_t \mid x)$ by a local quadratic model.
- With everything quadratic, obtain closed-form recursion for μ_t, P_t.

Local quadratic model

$$-\log p(v) \;:=\; r(v) \;\approx\; \tfrac{1}{2}(v-m)^{\mathsf{T}} M (v-m) + \text{const}$$

- Locally approximate near the predicted residual:

$$\bar{v} = y_t - C\mu_{t|t-1}$$

- Choose M and m so that:
  - m is the *mode* $\arg\max_v p(v)$
  - M is diagonal and the gradient $\nabla r(\bar{v})$ matches approximation $M(\bar{v}-m)$
  - If $p(v)$ has bounded support, use ε-margin projection of $\bar{v}$

Optimization interpretation

MAP estimate of posterior:

$$\min_{\xi} \left\{ \underbrace{\tfrac{1}{2}(\xi - \mu_{t|t-1})^\mathsf{T} P_{t|t-1}^{-1} (\xi - \mu_{t|t-1})}_{\text{prior}} + \underbrace{r(y_t - C\xi)}_{\text{measurement}} \right\},$$

- Our method (quadratic approximation of $r(y_t - C\xi)$ about the point $\xi = \mu_{t|t-1}$) is similar to a Newton step
- Becomes exact Newton step if we use $M = \nabla^2 r(\bar{v})$, but performs poorly in bimodal/heavy-tail cases

Proposed estimator

Time update (same as KF):

$$\begin{aligned} P_{t|t-1} &= \Sigma_w + A P_{t-1} A^\mathsf{T} \\ \mu_{t|t-1} &= A \mu_{t-1} \end{aligned}$$

Measurement update:

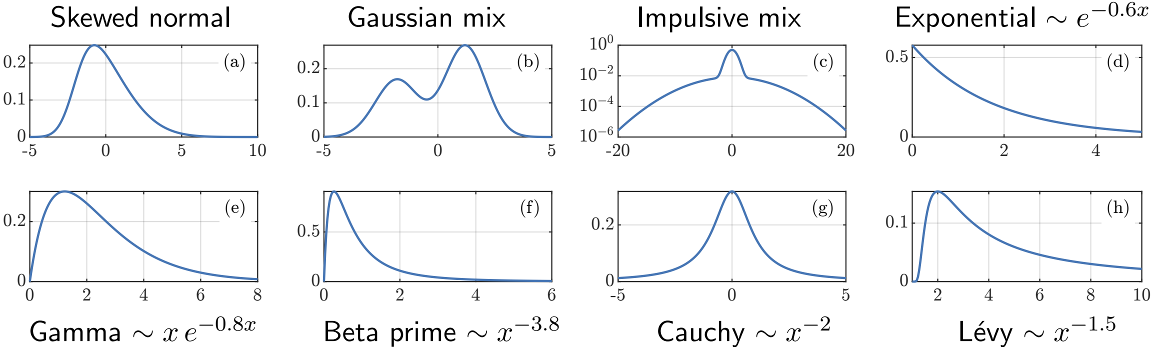
$$\begin{aligned} P_t^{-1} &= P_{t|t-1}^{-1} + C^\mathsf{T} M(\bar{v})\, C \\ \mu_t &= \mu_{t|t-1} + P_t\, C^\mathsf{T} \nabla r(\bar{v}) \end{aligned}$$

Properties:

- Estimator is *nonlinear*
- Gaussian case: $M(\bar{v}) = \Sigma_v^{-1}$ and $\nabla r(\bar{v}) = \Sigma_v^{-1}(\bar{v} - \mu_v)$, and we recover KF.
- Very efficient: No Monte Carlo, no optimization.

Measurement noise models

- Tuned to comparable mean/variance when possible.
- Ordered by heaviness of tails / boundedness of support.



Simulation setup

2D rotational system:

$$\begin{cases} x_{t+1} = \begin{bmatrix} \cos\frac{\pi}{18} & -\sin\frac{\pi}{18} \\ \sin\frac{\pi}{18} & \cos\frac{\pi}{18} \end{bmatrix} x_t + w_t \\ y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} x_t + v_t \end{cases}$$

- $w_t \sim \mathcal{N}(0, 0.05 I_2)$ and v_t non-Gaussian

Simulation setup

2D rotational system:
$$\begin{cases} x_{t+1} = \begin{bmatrix} \cos\frac{\pi}{18} & -\sin\frac{\pi}{18} \\ \sin\frac{\pi}{18} & \cos\frac{\pi}{18} \end{bmatrix} x_t + w_t \\ y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} x_t + v_t \end{cases}$$

- $w_t \sim \mathcal{N}(0, 0.05 I_2)$ and v_t non-Gaussian
- Compare our estimator to:
    - Maximum Correntropy KF (MCKF)[2]
    - Masreliez filter[3]
    - Kalman filter (KF) and particle filter (PF) baselines

[2] B. Chen, X. Liu, H. Zhao, and J. C. Príncipe, "Maximum correntropy Kalman filter," Automatica (2017).

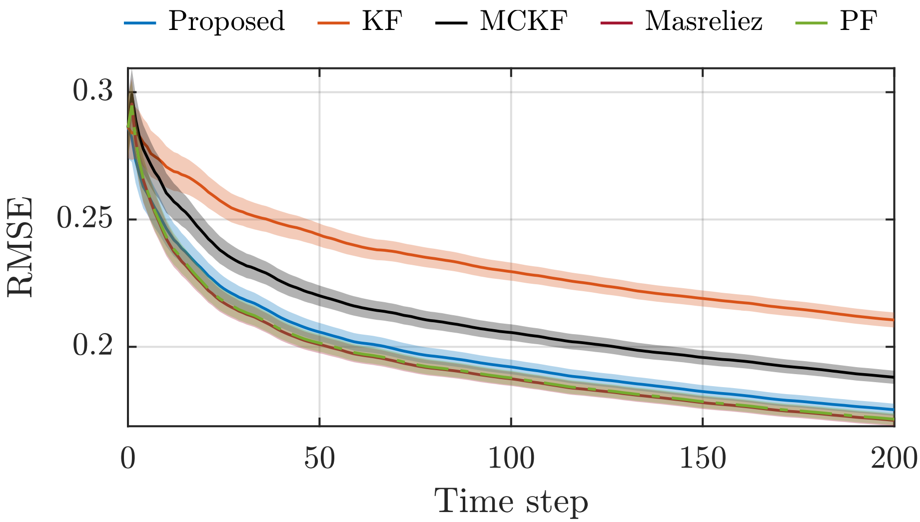
[3] C. J. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," TAC (1975).

Comparison of RMSE (200 trials)

Noise	KF	MCKF	Masreliez	Ours	PF
(a) Skewed normal	0.213	0.212	0.206	**0.205**	0.205
(b) Bimodal Gaussian mix	0.217	0.318	**0.202**	0.209	0.201
(c) Impulsive mix	0.212	0.191	**0.173**	0.177	0.173
(d) Exponential	0.212	0.206	∞	**0.203**	0.177
(e) Gamma	0.213	0.212	∞	**0.198**	0.191
(f) Beta prime	0.204	0.201	∞	**0.184**	0.156
(g) Cauchy	—	—	**0.216**	0.228	0.216
(h) Lévy	—	—	∞	**0.254**	0.242
Normalized time	1.00	7.25	40.5	1.05	69.3

Good performance on RMSE even though we approximate MAP!

Example: Impulsive Gaussian mixture



Take-home messages

- MAP estimation satisfies a DP recursion.
- Approximate dynamic programming leads to new estimators.
- Proposed: Local quadratic model for measurement log-density.
- Same structure and complexity as KF; reduces to KF when noise is Gaussian.
- Near particle-filter performance on many non-Gaussian noise models.

Code available:

Thank you!