Aalto University
School of Electrical Engineering

Updates in Bayesian filtering by Continuous Projections on a Manifold of Densities

Filip Tronarp & Simo Särkkä
Department of Electrical Engineering and Automation, Aalto University, Finland

Problem formulation

Continuous-discrete filtering problem:

$$\mathrm{d}X(t) = \mu(X(t),t)\,\mathrm{d}t + \sigma(X(t),t)\,\mathrm{d}B(t),$$
$$Y(t_n) \mid X(t_n) \sim f(y(t_n) \mid X(t_n)).$$

How to approximate Bayes' rule?

$$p(x(t_n) \mid \mathscr{Y}(t_n)) \propto f(y(t_n) \mid x(t_n)) p(x(t_n) \mid \mathscr{Y}(t_n^-))$$

Some Information Geometry

- $\mathscr{P}$ be a set of probability densities on $\mathscr{X} \subset \mathbb{R}^d$ then $p^{1/2}$ belongs to a sub-manifold of $\mathscr{L}^2$ (unit sphere).
- For a parametric susbset, $\mathscr{P}_\Theta \subset \mathscr{P}$, then p_θ, $\theta \in \Theta \subset \mathbb{R}^m$ belongs to a sub-manifold of the unit sphere.
- If $v = \frac{1}{2} p_\theta^{1/2} u \in \mathscr{L}^2$, then its projection onto the tangent space at θ is given by [1, Lemma 2.1]

$$\Pi_\theta v = \frac{1}{2}\mathbb{E}_\theta\left[u \nabla_\theta^T \log p_\theta\right] g^{-1}(\theta)(\nabla_\theta \log p_\theta) p_\theta^{1/2}, \quad (3)$$

  where $g(\theta)$ is the Fisher information matrix.
- In general the projection of

$$\partial_\tau p_\tau^{1/2} = \mathscr{A}(p_\tau^{1/2}), \quad p_0^{1/2} = p_{\theta_0}^{1/2} \in \mathscr{L}_\Theta^2,$$

  onto the parametric sub-manifold is given by [1]

$$\partial_\tau \hat{p}_{\theta(\tau)}^{1/2} = \Pi_{\theta(\tau)} \circ \mathscr{A}(\hat{p}_{\theta(\tau)}^{1/2}), \quad \hat{p}_{\theta(0)}^{1/2} = p_{\theta_0}^{1/2}. \quad (4)$$

The Projection Update (PU)

- Let $\pi_{\theta_0} \in \mathscr{P}_\Theta$ be some prior and $f(y \mid x) = \exp(\ell(x))$ be some likelihood then the posterior is

$$\pi(x \mid y) = \frac{f(y \mid x)\pi_{\theta_0}(x)}{\int_{\mathscr{X}} f(y \mid x)\pi_{\theta_0}(x)\,\mathrm{d}x}.$$

- Differentiable mapping from $\pi_{\theta_0}^{1/2}(x)$ to $\pi^{1/2}(x \mid y)$:

$$\mathscr{A}_{X|Y}(u) = \frac{1}{2}(\ell(x) - \mathbb{E}_{u^2}[\ell(X)])u, \quad u \in \mathscr{L}^2, \quad (5a)$$
$$\partial_\tau p_\tau^{1/2}(x \mid y) = \mathscr{A}_{X|Y}(p_\tau^{1/2}(x \mid y)), \ \tau \in [0,1]. \quad (5b)$$

- The projection update is defined by

$$\partial_\tau \hat{p}_{\theta(\tau)}^{1/2} = \Pi_{\theta(\tau)} \circ \mathscr{A}_{X|Y}(\hat{p}_{\theta(\tau)}^{1/2}), \quad \hat{p}_{\theta(0)}^{1/2} = \pi_{\theta_0}^{1/2}. \quad (6)$$

  – $\mathscr{P}_\Theta$ is an exponential family with sufficient statistic $T(X)$:

$$\partial_\tau \theta(\tau) = [g(\theta(\tau))]^{-1} \widehat{\mathbb{C}}_\tau[T(X), \ell(X)]. \quad (7)$$

  – $\mathscr{P}_\Theta$ is the Gaussian family:

$$\partial_\tau \mu = \widehat{\mathbb{E}}_\tau[(X-\mu)\ell(X)], \quad (8a)$$
$$\partial_\tau \Sigma = \widehat{\mathbb{E}}_\tau[(X-\mu)(X-\mu)^\mathsf{T}(\ell(X) - \mathbb{E}[\ell(X)])]. \quad (8b)$$

- Exact updates when an exponential (including Gaussian) family $\mathscr{P}_\Theta$ is conjugate prior to the likelihood (Theorem 1 & 2 in the paper).

Conclusions

- A curve from prior to posterior can be defined and projected onto a manifold of densities, giving rise to *projection updates*.
- The method provides an effective way to for Bayesian updates with non-Gaussian likelihoods.

Results I: Wiener Velocity with Outlier Measurements

$X(t)$ follows a Wiener velocity model and is measured by

$$Y(t_n) = [0, \mathrm{I}_2]X(t_n) + R_n^{1/2} V_n.$$

where $R_n = R_0 = \mathrm{I}$ with probability $1-\alpha$ and $R_n = 20R_0$. As in [2] (MM) the noise is modelled with a Laplace distribution and the projection update (PU) is compared with (MM) and a Kalman filter (KF) using $R_n = R_0$. RMSE for position is shown in Figure 1.

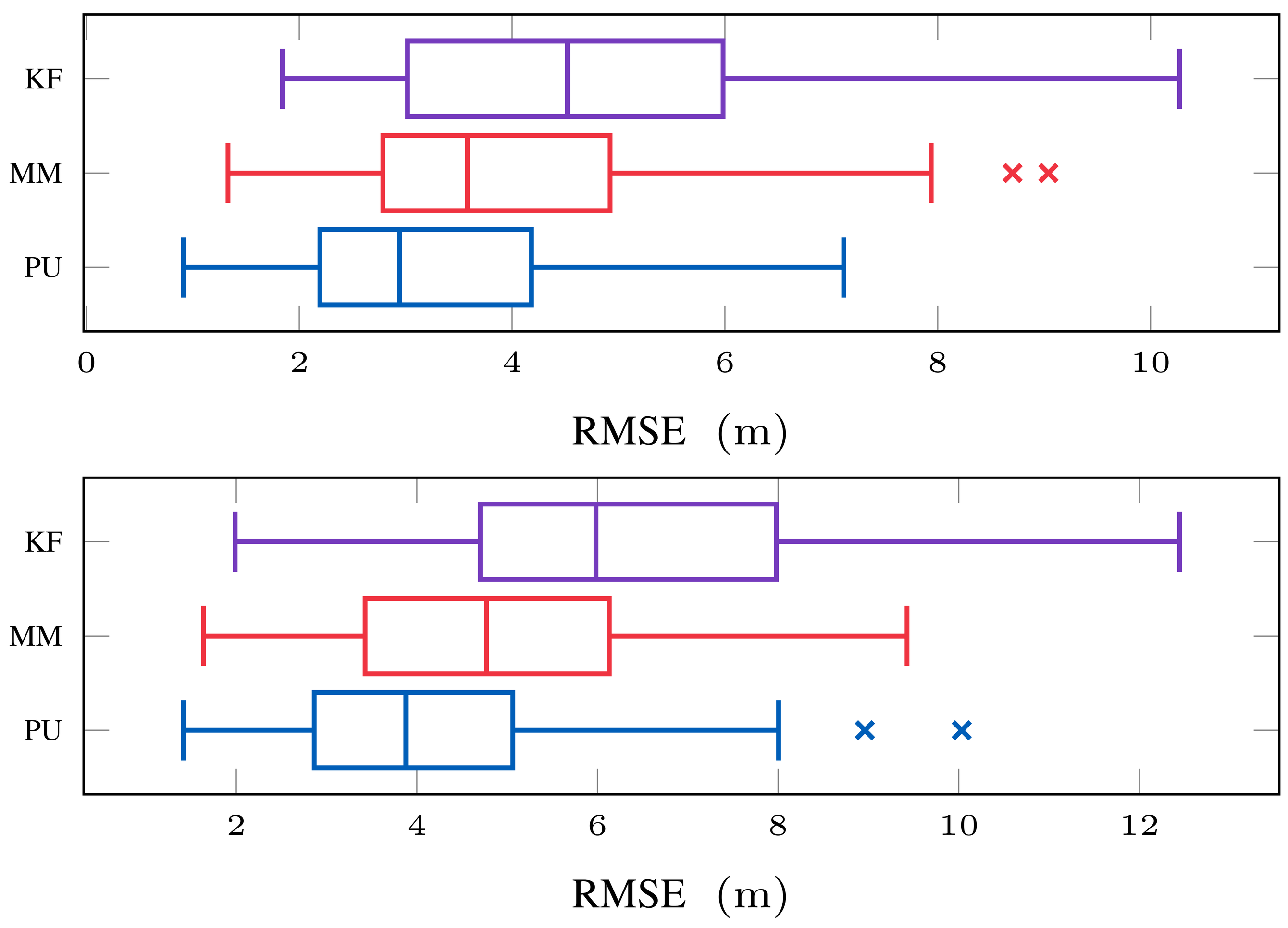

Figure 1. $\alpha = 0.2/0.4$ for figure above/below.

Results II: Stochastic Volatility

$$\mathrm{d}X(t) = -\lambda(X(t) - m)\,\mathrm{d}t + \sigma B(t),$$
$$Y(t_n) = \exp(X(t_n)/2)V_n, \ V_n \sim \mathscr{N}(0,1).$$

The initial condition is Gaussian with moments $\mathbb{E}[X(0)] = \mathbb{V}[X(0)] = 1$ and $\sigma = m = 1$. We compare PU to Laplace approximation (LA) and a Kalman filter (KF) with transformed measurements (see [3]). RMSE is shown in Figure 2.

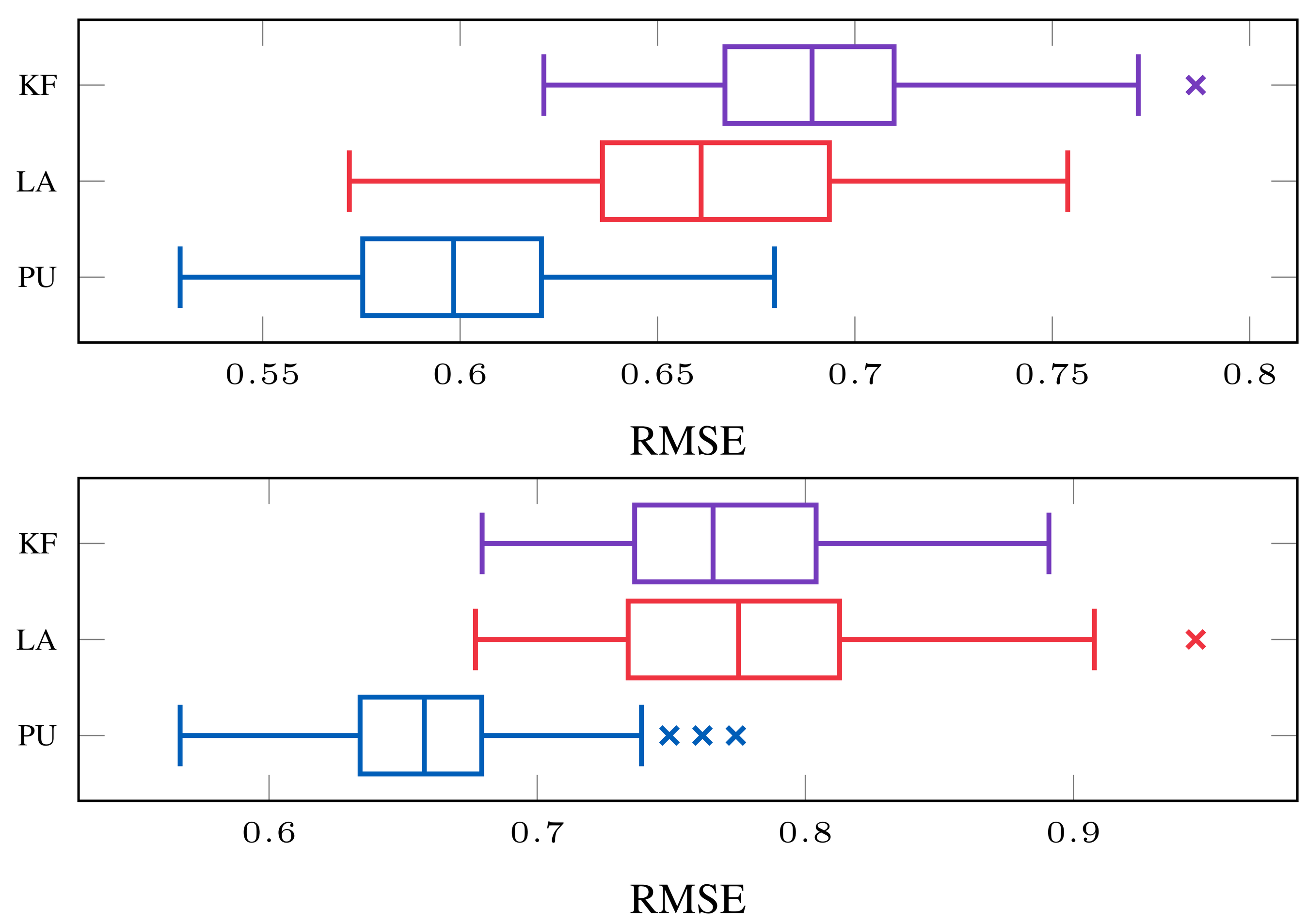

Figure 2. $\lambda = 0.5/0.1$ for figure above/below.

References

[1] D. Brigo, B. Hanzon, and F. LeGland, "Approximate nonlinear filtering by projection on exponential manifolds of densities," *Bernoulli*, vol. 5, no. 3, pp. 495–532, 1999.

[2] H. Wang, H. Li, W. Zhang, and H. Wang, "Laplace ℓ_1 robust Kalman filter based on majorization minimization," *in 20th International Conference on Information Fusion.*

[3] K. Platabioti, EJ McCoy, and DA Stephens, "A review of stochastic volatility: univariate and multivariate models," Tech Rep., 2005.

Department of Electrical Engineering and Automation
School of Electrical Engineering
Aalto University, Finland

Contact information for comments and questions:
Filip Tronarp (email: filip.tronarp@aalto.fi)