Multimodal Nonlinear Filtering Using Gauss-Hermite Quadrature

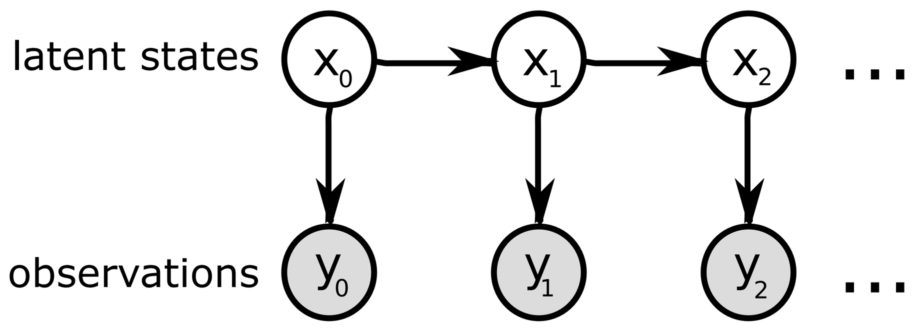
Hannes P Saal, Nicolas M O Heess, and Sethu Vijayakumar
School of Informatics, University of Edinburgh

Background

In **filtering problems** the (posterior) state distribution $p(\mathbf{x})$ is recursively estimated given observations $\mathbf{y}$ and state dynamics. For **nonlinear** observation functions, the state distribution can become **multimodal**. Common Solutions are

1. Unimodal Gaussian filter using linearization of obs. function.
2. "Bank of filters" using multiple independent unimodal filters.

We present a **variational approach** for recusively fitting a mixture of Gaussians (MoG) to state posteriors that minimizes the **KL divergence** between a MoG and the true posterior state distribution.



Problem statement: Here, we only consider the observation update. Given the following prior, likelihood and observation function:

Prior: Mixture of Gaussians	$p(\mathbf{x}) = \sum_n \gamma_n \mathcal{N}(\mathbf{x}\vert\boldsymbol{\nu}_n, \boldsymbol{\Theta}_n)$
Likelihood: Gaussian	$p(\mathbf{y}\vert\mathbf{x}) = \mathcal{N}(\mathbf{y}\vert\mathbf{f}(\mathbf{x}), \boldsymbol{\Sigma}_y)$
Obs. function: RBF network	$f(\mathbf{x}) = \sum_j c_j k(\mathbf{x}, \mathbf{m}_j)$

where $k(\mathbf{x}, \mathbf{m}_j) = \exp\left\{-0.5(\mathbf{x}-\mathbf{m}_j)^T \mathbf{S}^{-1}(\mathbf{x}-\mathbf{m}_j)\right\}$

We recursively approximate the posterior with a MoG $q_{\text{mix}}(\mathbf{x}) = \sum_m \alpha_m \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$:

$$p(\mathbf{x}|\mathbf{y}_t, \mathbf{y}_{t-1}, ..., \mathbf{y}_0) \propto p(\mathbf{y}_t|\mathbf{x})p(\mathbf{x}|\mathbf{y}_{t-1}, ..., \mathbf{y}_0)$$

Variational approach using deterministic sampling

Idea: Minimize KL divergence between MoG approximation and true posterior:

$$\mathrm{KL}\left[q_{\text{mix}}||p\right] = \int \mathrm{d}\mathbf{x}\; q_{\text{mix}}(\mathbf{x}) \log \frac{q_{\text{mix}}(\mathbf{x})}{p(\mathbf{x}|\mathbf{y})}$$
$$= -\underbrace{\mathrm{H}\left[q_{\text{mix}}\right]}_{\text{intractable}} - \sum_m \alpha_m \underbrace{\mathrm{E}_{q_m}\left[\log p(\mathbf{x})\right]}_{\text{intractable}} - \sum_m \alpha_m \mathrm{E}_{q_m}\left[\log p(\mathbf{y}|\mathbf{x})\right] + C$$

Problem: Integrals over log-sums in the KL divergence are intractable.
Solution: We use Gauss-Hermite quadrature (deterministic sampling) to approximate such integrals:

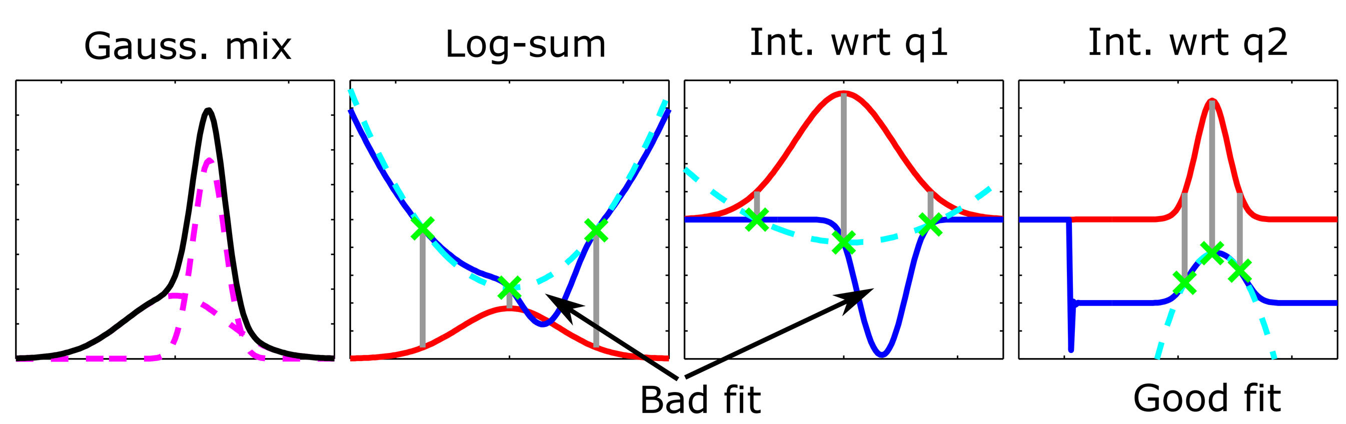
$$\int \mathrm{d}\mathbf{x} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) f(\mathbf{x}) = \mathrm{E}_{q_m}\left[f(\mathbf{x})\right] \approx \pi^{-\frac{d}{2}} \sum_{\mathbf{h}} w_{\mathbf{h}} f(\mathbf{z}_{\mathbf{h}})$$

To improve the quality of the approximation in situations where the covariances of the q_i differ considerably, we rewrite expectations over log-sums as a sum:

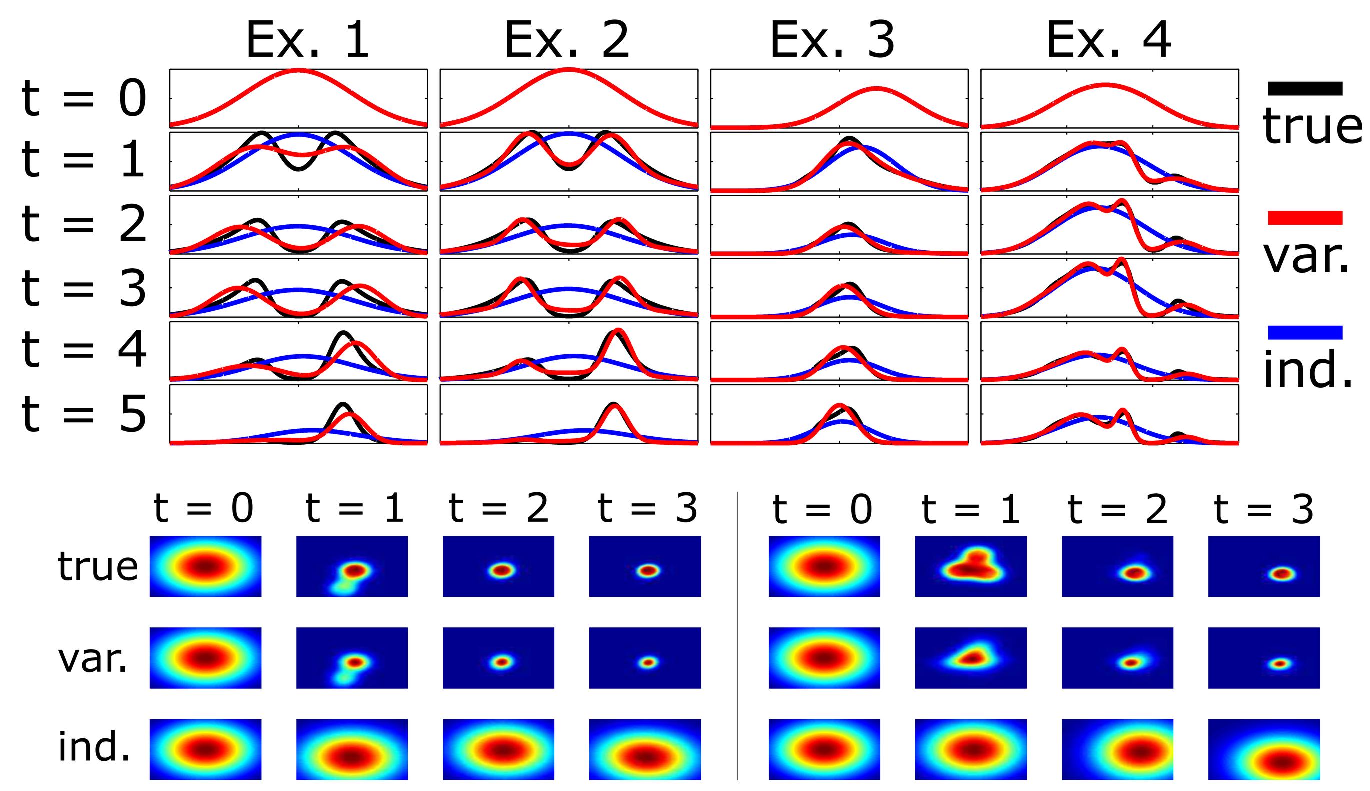
$$\mathrm{E}_{q_m}\left[\log \sum_n \alpha_n \mathcal{N}_n\right] = \mathrm{E}_{q_m}\left[\log \alpha_1 \mathcal{N}_1\right] + \sum_{n=2}^{N} \mathrm{E}_{q_m}\left[\log\left(1 + \frac{\alpha_n \mathcal{N}_n}{\sum_{k=1}^{n-1} \alpha_k \mathcal{N}_k}\right)\right]$$

and then integrate over individual components separately

$$\underbrace{\mathrm{E}_{q_m}\left[\log\left(1 + \frac{\alpha_n \mathcal{N}_n}{\sum_{k=1}^{n-1} \alpha_k \mathcal{N}_k}\right)\right]}_{\text{Integral over } q_m} = \underbrace{\mathrm{E}_{q_n}\left[\frac{\mathcal{N}_m}{\mathcal{N}_n} \log\left(1 + \frac{\alpha_n \mathcal{N}_n}{\sum_{k=1}^{n-1} \alpha_k \mathcal{N}_k}\right)\right]}_{\text{Integral over } q_n}$$

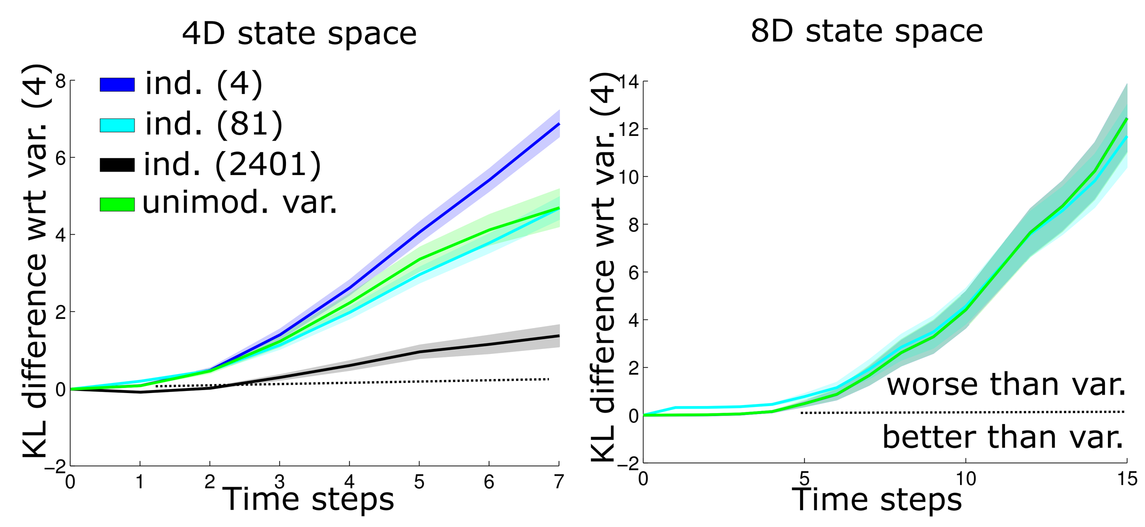


Qualitative results



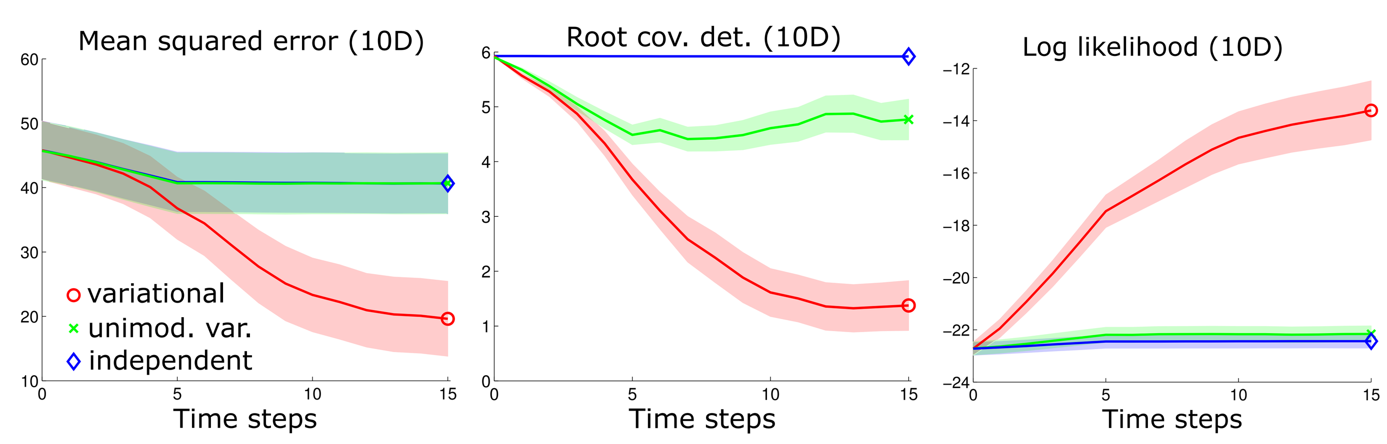
Quantitative evaluation of posterior approx.

We compared the fit of various mixture approximations with the true posterior by calculating differences of **KL divergences** $\mathrm{KL}\left[q||p\right] - \mathrm{KL}\left[q_{mix}||p\right]$ for alternative approximations q. The "bank of filters" approach fits the posterior considerably worse, even when a high number of components is used.

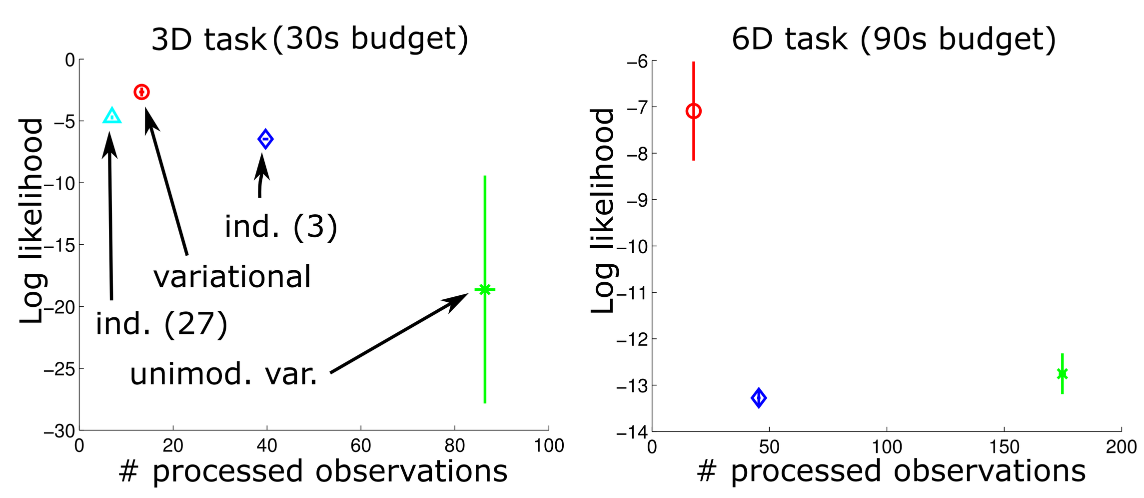


Convergence on active learning task

We compared performance on a localization task with a static target x^* and obs. function $f(\mathbf{x}-\boldsymbol{\theta})$. Query points $\boldsymbol{\theta}_t$ were chosen so as to minimize posterior uncertainty (**active learning**). For such problems, an accurate representation of the posterior uncertainty is important.We found that the variational method converges while other methods stall on this problem.



Even though the variational method is computationally more expensive and processes fewer observations than other methods per unit of time, its performance is still better, when a fixed **time budget** is imposed.



Conclusions

- Improved quality of approximation of variational MoG distribution over independent mixture filters.
- Efficient calculation of intractable integrals using deterministic sampling.
- Tested on problems up to 10D.

Acknowledgements: This study is supported by the EPSRC/MRC and the EU-funded Sensopac project. Contact: Hannes Saal (hannes.saal@gmail.com).