Approximate Bayesian Inference Team

Mohammad Emtiyaz Khan

Goals and Challenges

Goal: To design AI that can continually learn using Bayesian uncertainty.

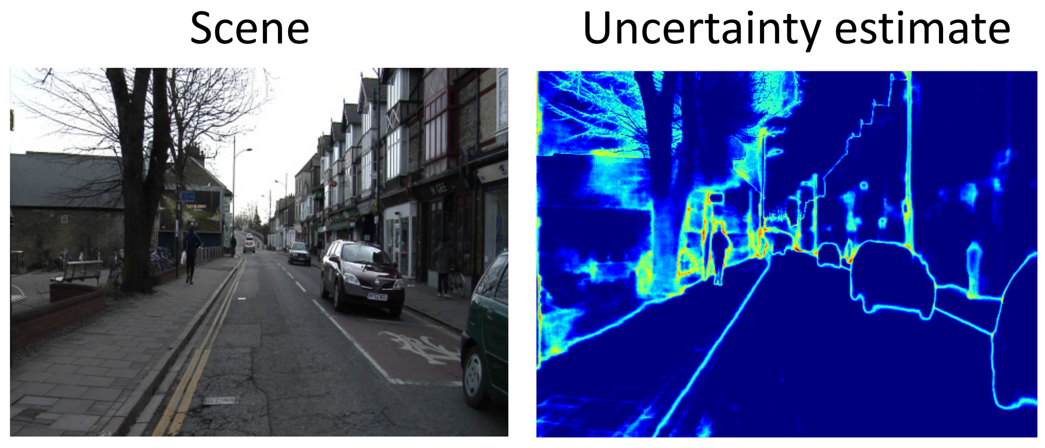
Examples: Uncertainty, i.e., knowing how much we don't know, is useful in designing:

- Robots that can understand their environments (e.g., for elderly care)
- AI to design high-performance buildings
- Methods that improve performance of deep-learning methods

Challenge: Computation of Bayesian uncertainty is difficult

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\underbrace{\int p(\mathcal{D}|\theta)p(\theta)d\theta}_{\text{Intractable integral}}}$$

Main Idea: Approximate integration by optimization



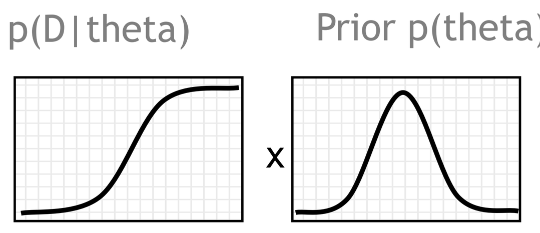
For depth estimation, the algorithm is uncertain about the edges in the scene (see right image). This knowledge can improve safety for self-driving cars (taken from Kendall et. al. 2017)

Simplifying Approximate Bayesian Inference for Complex Models

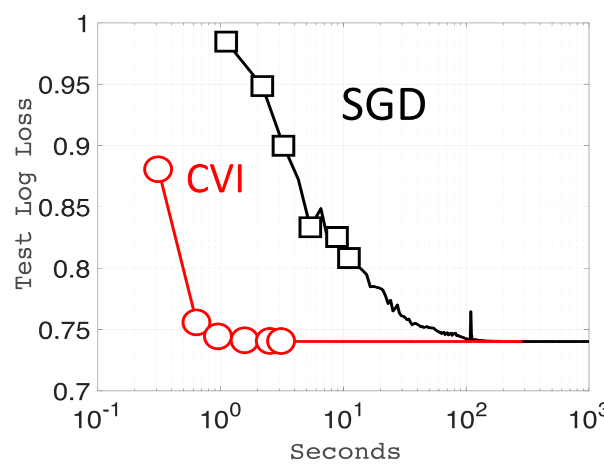
For non-conjugate models, we propose Conjugate-Computation Variational Inference (CVI) to approximate the Bayesian integration

- CVI is a general method that is also easy to implement for many models
- It expresses inference in a complex model as an inference in a simpler model, e.g., logistic Regression expressed as linear Regression.
- For a general graphical model, this can be implemented using message passing

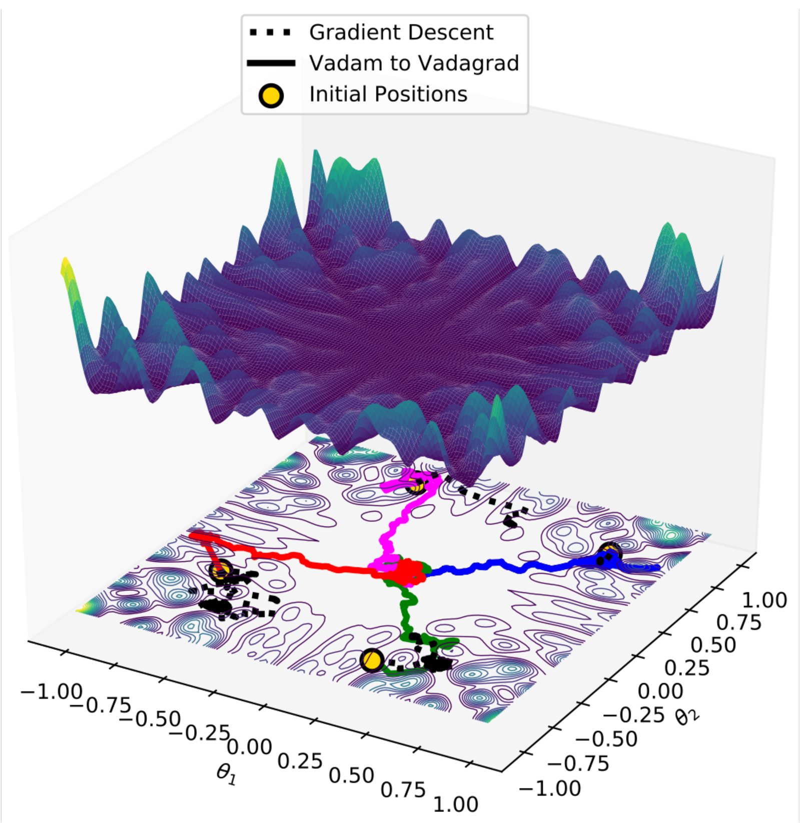
Bayesian Integral is difficult



Logistic Regression
(N = 581K, D = 54)

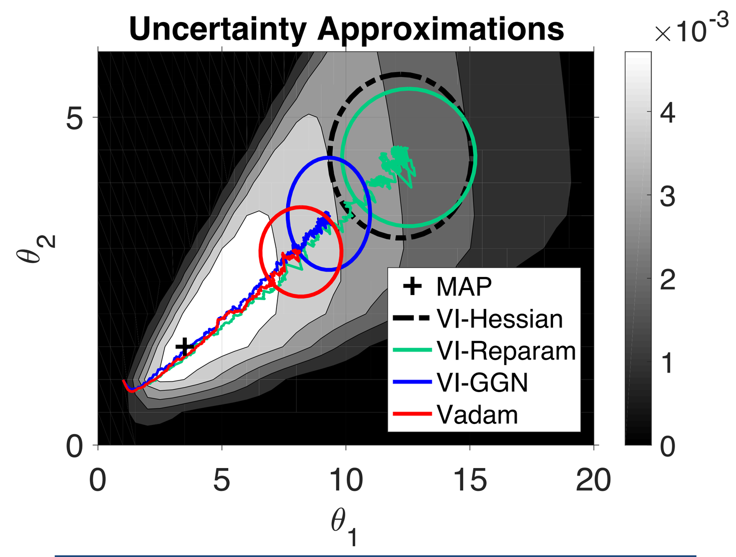


Avoiding Local Minima



Our method (thick lines) reaches flat minima, while gradient descent (black lines) gets stuck.

Quality of Uncertainty Estimates



For 2-D logistic, variances for our method (red line) are shrunk compared to the truth (dashed black lines)

Fast Inference for Deep Learning

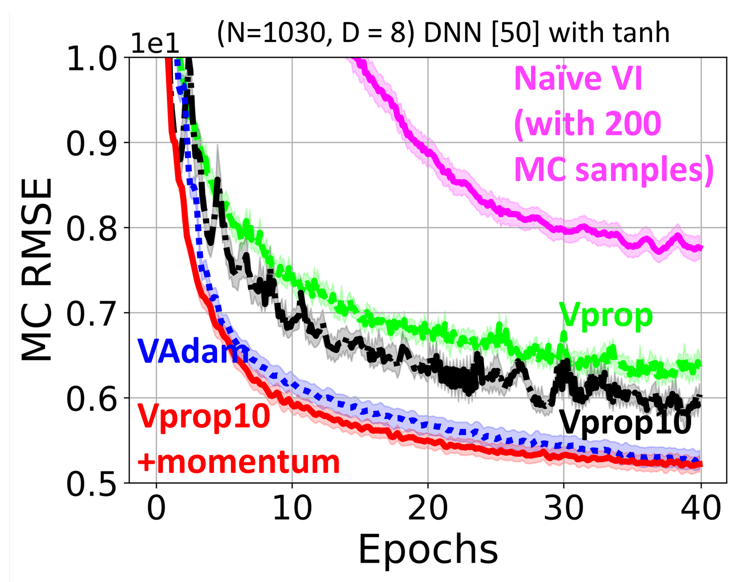
To enable Gaussian variational inference in deep models, we slightly modify existing deep learning methods, such as RMSprop and Adam, to perform variational inference

- Perturb the network weights during gradient evaluations
- Requires a few lines of code change in the existing software
- Scales to very large problems with minimal effort

Our method also improves existing optimizers: reduce overfitting, avoid local minima, and encourages exploration.

RMSprop for Maximum Likelihood Estimation	Vprop for Approximate Bayesian Inference
$\theta \leftarrow \mu$	$\theta \leftarrow \mu + \epsilon/\sqrt{s+\lambda}$
$g \leftarrow \nabla_\theta f(\theta)$	$g \leftarrow \nabla_\theta f(\theta) + \lambda\mu$
$s \leftarrow (1-\beta)s + \beta g^2$	$s \leftarrow (1-\beta)s + \beta g^2$
$\mu \leftarrow \mu - \alpha \frac{g}{\sqrt{s}+\delta}$	$\mu \leftarrow \mu - \alpha \frac{g}{\sqrt{s}+\lambda}$

Bayesian Regression with DNN

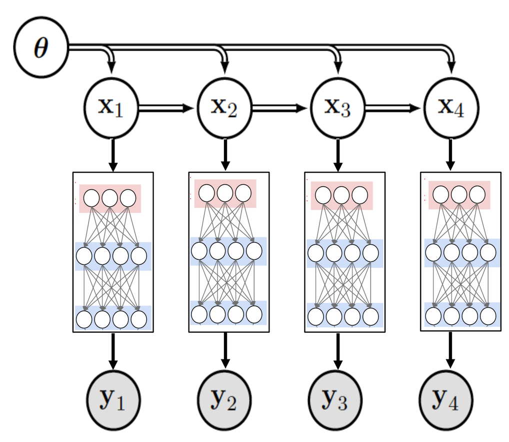


Our method (in red) is much faster than the existing methods (in magenta) and also require fewer samples

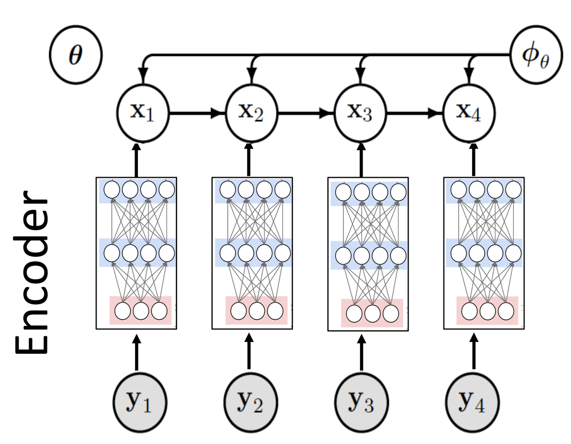
Scalable Inference in Structured Deep Models Using Message Passing

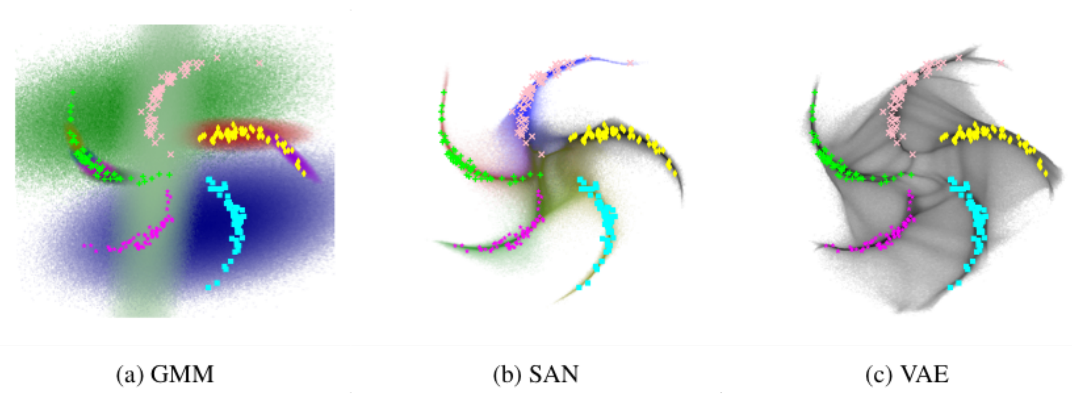
For Structured deep models (shown in the right), we propose an efficient message passing algorithm. Our algorithm is a structured, amortized, and natural-gradient (SAN) inference algorithm.

Graphical model + Deep Models



Structured Inference Networks





Our SAN method is as flexible as VAE but can also cluster the data similar to GMM

References

1. Conjugate-Computation Variational Inference, (AISTATS 2017) M.E. KHAN AND W. LIN
2. Vprop: Variational Inference using RMSprop, (NIPS 2017, WORKSHOP ON BAYESIAN DEEP LEARNING), M.E. KHAN, Z. LIU, V. TANGKARATT, AND Y. GAL
3. Variational Adaptive-Newton Method for Explorative-Learning, (NIPS 2017, WORKSHOP ONAABI) M.E. KHAN, W. LIN, V. TANGKARATT, Z. LIU, AND D. NIELSEN
4. Variational Message Passing with Structured Inference Networks, (ICLR 2018) W. LIN, N. HUBACHER, AND M.E. KHAN