Reparameterization Gradient Message Passing

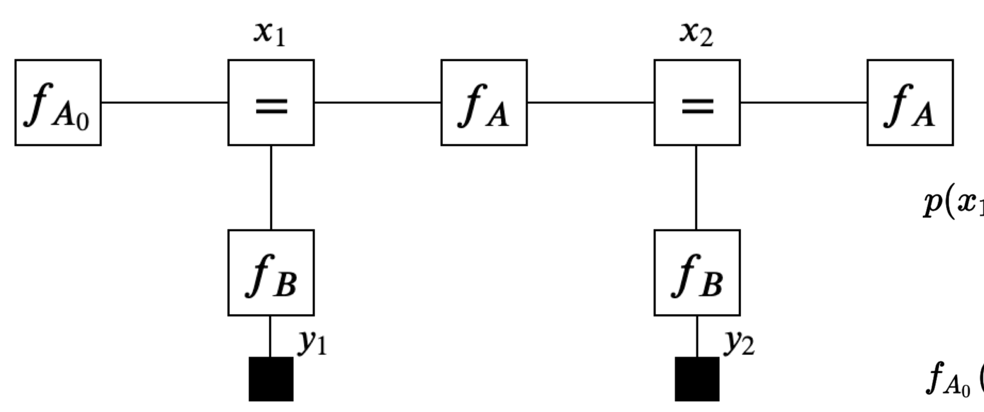
Technische Universiteit Eindhoven University of Technology

BIASlab

Semih Akbayrak[1] and Bert de Vries[1,2]
[1]Eindhoven University of Technology, [2]GN Hearing, Eindhoven, The Netherlands
s.akbayrak@tue.nl, bdevries@ieee.org

Introduction and Motivation

- Message passing is an algorithmic approach to Bayesian inference problems and can be better understood within a factor graph framework.
- A Forney-style factor graph (FFG) is a graphical depiction of the independency structure in a probabilistic model where variables are represented with edges and nodes correspond to factors. In an FFG, variables can branch to more than two factors via "equality nodes" since an edge maximally connects two nodes.



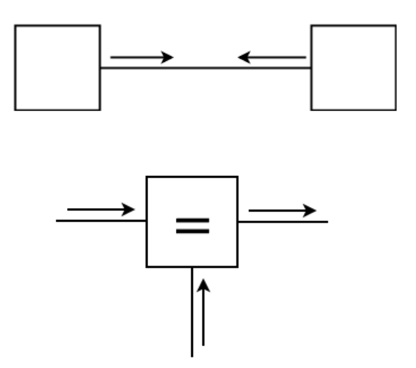
$$p(x_{1:T}, y_{1:T}) = p(x_1) \cdot p(y_1|x_1) \cdot \prod_{t=2}^{T} p(x_t|x_{t-1}) \cdot p(y_t|x_t)$$

$$f_{A_0}(x_1) = p(x_1), \;\; f_A(x_t, x_{t-1}) = p(x_t|x_{t-1})$$

$$f_B(y_t, x_t) = p(y_t|x_t)$$

- Existing message passing algorithms such as belief propagation (BP) and variational message passing (VMP) requires conditionally conjugate model structure so that multiplication of incoming messages leads to closed-form marginals.

Two common message multiplication types



- Marginal posterior of a random variable
- Iterative updates of VMP
- Update step in state space models

- Message passing based probabilistic programming libraries, e.g. ForneyLab [1], employ pre-defined message passing rules to execute inference by taking advantage of conjugacy.
- Message passing procedures may be interrupted if the model contains a non-conjugate part or a rule is not defined for the incoming message types.
- We use recent advances in stochastic variational inference (SVI) to extend the class of models for which inference by message passing can be performed.

A Quick Comparison of SVI and Message Passing

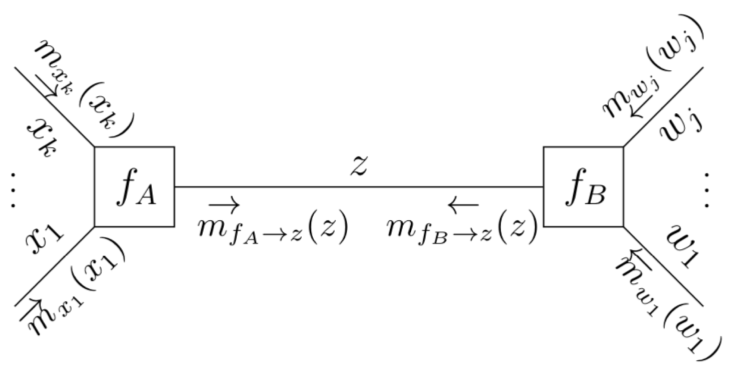
SVI: Stochastic optimization of a variational objective by using its noisy gradients.

Message Passing	Stochastic Variational Inference
▪ Fast	▪ Slow
▪ Hyperparameter-free	▪ Requires manually specified hyperparameters
▪ Exact for Belief Propagation	▪ Converges to local minima
▪ Limited to conditionally conjugate models	▪ Extends to broader range of models

- We combine the strengths of both approaches in an inference procedure called *Reparameterization Gradient Message Passing (RGMP)* that is faster than pure SVI and more general than BP and VMP.

Reparameterization Gradient Message Passing

- In an FFG, incoming messages to a variable (z) summarizes the rest of the graph, providing a prior and a likelihood function to the variable. This enables approximating the marginal by minimizing a local free energy.



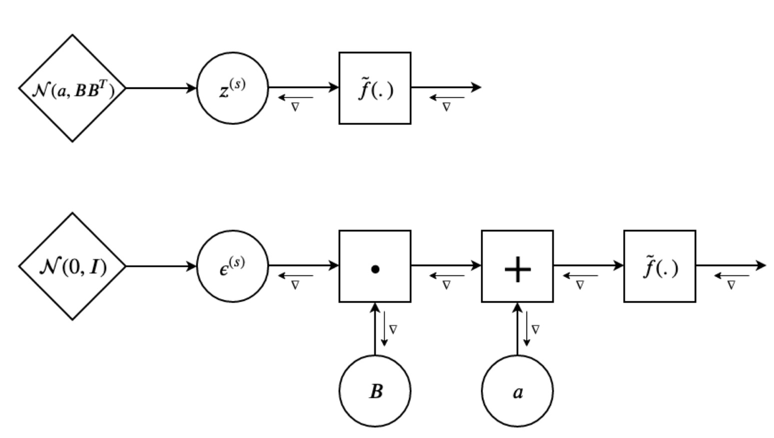
$$F(\phi) \triangleq \mathbb{E}_q\left[\log \frac{q(z;\phi)}{\tilde{f}(z)}\right] = \int q(z;\phi) \log \frac{q(z;\phi)}{\tilde{f}(z)} \, \mathrm{d}z$$

where $\tilde{f}(z) = m_{f_A \to z}(z) \cdot m_{f_B \to z}(z)$.

For most models it is difficult to estimate optimal variational parameters (ϕ), analytically. Additionally, it is often not trivial to evaluate the gradient of the free energy ($\nabla_\phi F$).

- Noisy gradients, which are computed with samples $z^{(s)} \sim q(z;\phi)$, lose some information of ϕ since $\nabla_\phi \log \tilde{f}(z^{(s)})$ becomes zero although $\tilde{f}(z^{(s)})$ strictly depends on ϕ.

- The reparameterization trick [2,3,4] addresses this problem by generating samples ($z^{(s)}$) from a differentiable process of dummy random variables $\epsilon^{(s)}$. Gaussian ex. is at right.
- The gradient of free energy can be expressed as expectation of the gradient.



$$\nabla_\phi F(\phi) = \mathbb{E}_{p_\epsilon(\epsilon)}\left[\nabla_\phi \log \frac{p_\epsilon(\epsilon)\left|\frac{\partial \epsilon}{\partial z}\right|}{\tilde{f}(g(\epsilon;\phi))}\right]$$

- The variational parameters can be iteratively updated by employing the noisy gradients within a stochastic optimization process.

The resulting approximation interfaces with BP and VMP by providing the posterior marginal of the variable connected to non-conjugate factor pairs.

Experimental Validation

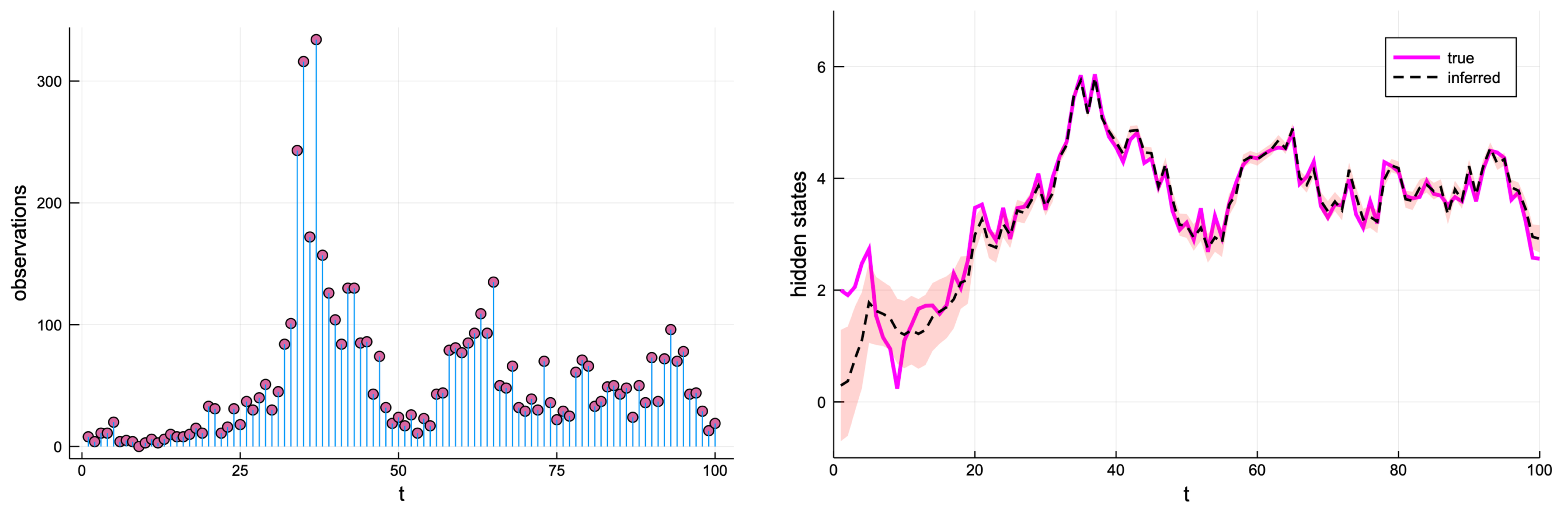
$$p(x_{1:T}, y_{1:T}) = p(x_1)p(y_1|x_1)\prod_{t=2}^{T} p(x_t|x_{t-1})p(y_t|x_t)$$

$$p(x_1) = N(x_1; 0, 1), \;\; p(x_t|x_{t-1}) = N(x_t; x_{t-1}, 0.2)$$

$$p(y_t|x_t) = \mathcal{P}o(y_t; \exp(x_t))$$

Poisson Linear Dynamical System (PLDS) Model Specification
Given $y_{1:T}$, estimate $x_{1:T}$.

Prediction messages are analytically computed with BP. Update messages are approximated with RGMP.



References

1. Cox, Marco, Thijs van de Laar, and Bert de Vries. "ForneyLab. jl: Fast and flexible automated inference through message passing in Julia."
2. Titsias, Michalis, and Miguel Lázaro-Gredilla. "Doubly stochastic variational Bayes for non-conjugate inference." *International conference on machine learning*. 2014.
3. Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).
4. Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra. "Stochastic backpropagation and approximate inference in deep generative models." *arXiv preprint arXiv:1401.4082* (2014).