Fast and Simple Natural-Gradient Variational Inference with Mixture of Exponential-family Approximations

Wu Lin (UBC)

June 11, 2019

Joint work with Mohammad Emtiyaz Khan (AIP, RIKEN) and Mark Schmidt (UBC)

Variational Inference (VI)

VI approximates the posterior $p(\mathbf{z}|\mathcal{D}) \approx q(\mathbf{z}|\lambda_z)$ by maximizing the evidence lower bound:

$$\text{ELBO:} \quad \max_{\lambda_z} \; \mathcal{L}(\boldsymbol{\lambda}_z) := \mathbb{E}_q\Big[\overbrace{\log p(\underbrace{\mathcal{D}}_{\text{data}}, \mathbf{z})}^{\text{Probabilistic Model}} - \log q(\mathbf{z}|\boldsymbol{\lambda}_z)\Big]$$

where $q(\mathbf{z})$ is a tractable distribution parametrized by $\boldsymbol{\lambda}_z$.

ELBO Optimization

Block-box VI (BBVI):

$$\boldsymbol{\lambda}_z \leftarrow \boldsymbol{\lambda}_z + \beta \nabla_{\lambda_z} \mathcal{L}(\boldsymbol{\lambda}_z)$$

Natural-gradient VI (NGVI):

$$\boldsymbol{\lambda}_z \leftarrow \boldsymbol{\lambda}_z + \beta \overbrace{\mathbf{F}_z(\boldsymbol{\lambda}_z)^{-1} \nabla_{\lambda_z} \mathcal{L}(\boldsymbol{\lambda}_z)}^{\text{natural gradient}}$$

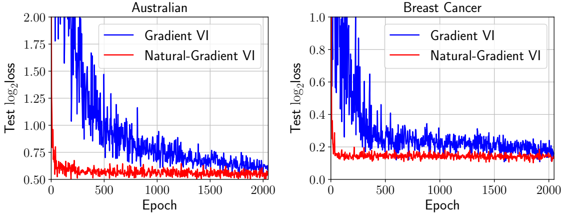
where $\mathbf{F}_z(\boldsymbol{\lambda}_z)$ is the Fisher information matrix of $q(\mathbf{z}|\boldsymbol{\lambda}_z)$.

Advantages of NGVI:

- NGVI can be simple and fast when q is in the exponential family (e.g., Gaussian) (Khan and Lin, AI&Stats 2017).

$$\text{NGVI for Exp-Family: } \boldsymbol{\lambda}_z \leftarrow \boldsymbol{\lambda}_z + \beta\, \nabla_{m_z} \mathcal{L}(\boldsymbol{\lambda}_z)$$

because $\nabla_{m_z} \mathcal{L}(\boldsymbol{\lambda}_z) = \mathbf{F}_z(\boldsymbol{\lambda}_z)^{-1} \nabla_{\lambda_z} \mathcal{L}(\boldsymbol{\lambda}_z)$.

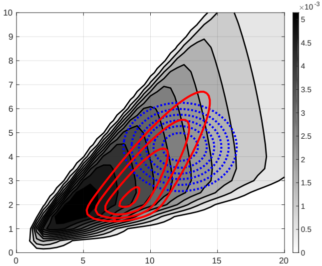


Problem Formulation

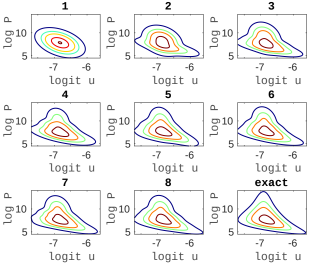
Challenges of NGVI when $q(\mathbf{z})$ is not in the exponential-family :

- Computing $\mathbf{F}_z(\boldsymbol{\lambda}_z)^{-1}\nabla_{\lambda_z}\mathcal{L}(\boldsymbol{\lambda}_z)$ could be complicated.
- $\mathbf{F}_z(\boldsymbol{\lambda}_z)$ can be singular.
- Often no simple update beyond exponential family.

Our goal: perform a simple NGVI update for more flexible variational approximations (e.g., skewness, multi-modality)



(a) Skew Gaussian



(b) Finite Mixture of Gaussians

This Work

Main Contribution: propose a new NGVI update for a class of mixture of exponential family distributions.

We consider the following mixture:

$$q(\mathbf{z}|\boldsymbol{\lambda}) = \int \underbrace{q(\mathbf{z}|\mathbf{w}, \boldsymbol{\lambda}_z)}_{\text{exp-family}} \underbrace{q(\mathbf{w}|\boldsymbol{\lambda}_w)}_{\text{exp-family}} d\mathbf{w}$$

We propose to use the (joint) Fisher matrix $\mathbf{F}_{wz}$ of $q(\mathbf{w}, \mathbf{z}|\boldsymbol{\lambda})$ since:

$$\nabla_m \mathcal{L}(\boldsymbol{\lambda}) = \mathbf{F}_{wz}(\boldsymbol{\lambda})^{-1} \nabla_\lambda \mathcal{L}(\boldsymbol{\lambda})$$

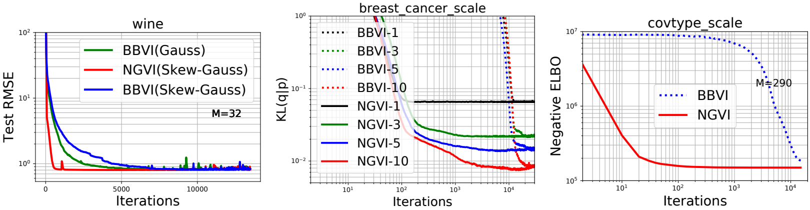
where $\mathbf{m}$ is the proposed expectation parameter.

- Proposed NGVI update: $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \beta \nabla_m \mathcal{L}(\boldsymbol{\lambda})$

Proposed NGVI

Advantage of the proposed NGVI:

- Has the same cost as BBVI if computing $\nabla_m \mathcal{L}(\boldsymbol{\lambda})$ is easy.
- Is faster than BBVI.



Variational approximations:

- Finite mixture of exp-family distributions:
  Mixture of Gaussians (multi-modality)
  Birnbaum-Saunders distribution (non-Gaussian mixture)
- Gaussian compound distribution:
  Skew Gaussian (skewness)
  Normal inverse-Gaussian (heavy tails)

Summary & Poster Presentation

Conclusion:
a simple NGVI update for approximations outside the exp-family.

Poster Presentation:

- This work:
  Poster #217, Pacific Ballroom, Today, 6:30 PM

- New gradient estimators via Stein's lemma:
  "Stein's Lemma for the Reparameterization Trick with Exponential-family Mixtures", the workshop on Stein's method, Saturday