

Boosting Black Box Variational Inference

Francesco Locatello^{1,2,3} *, Gideon Dresdner¹ *, Rajiv Khanna⁴,
Isabel Valera², Gunnar Rätsch¹

NeurIPS

[Dec 6, 2018]

ETH zürich



Max Planck ETH Center
for Learning Systems



ETH zürich

Berkeley
UNIVERSITY OF CALIFORNIA

* joint first author

Boosting Variational Inference

$$\text{Variational Inference} \quad \arg \min_{q \in \mathcal{Q}} D^{KL}(q(z)|p(z|x))$$

$$\text{Boosting Variational Inference} \quad \arg \min_{q \in \text{conv}(\mathcal{Q})} D^{KL}(q(z)|p(z|x))$$

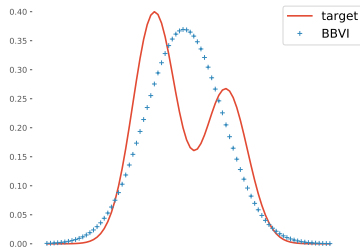
- Variational Inference

- ▷ Trade-off between the tractability of the problem and expressivity of \mathcal{Q}

- Boosting Variational Inference

- ▷ Bayesian Inference meets Convex-Optimization
- ▷ Allows for tradeoff between additional run-time for a better approximation

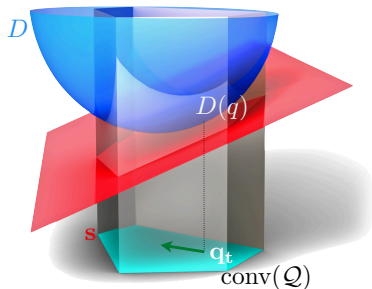
Locatello et al., 2018; Guo et al., 2017; Miller et al., 2016



Boosting Variational Inference with Frank-Wolfe

- Convergence rates $\mathcal{O}(1/t)$
- Several restrictive assumption for provable convergence for Variational Inference
- Knowledge of the optimizer to design linear minimization subroutine
- In this work: we relax assumptions and allow for black box solvers

$$\arg \min_{q \in \text{conv}(\mathcal{Q})} D^{KL}(q(z)|p(z|x))$$



Frank & Wolfe 1956; Locatello et al., 2018

Residual ELBO (RELBO)

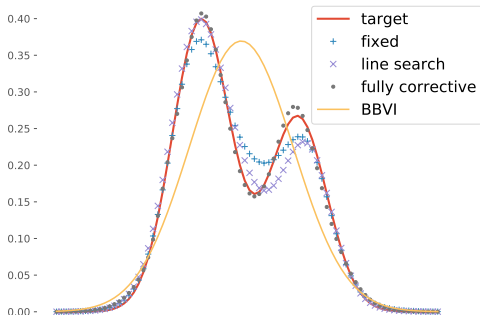
Inference as an iterative procedure

$$\text{Subroutine} \quad \arg \min_{s(z) \in \mathcal{Q}} D^{KL} \left(s(z) \parallel \sqrt[\lambda]{\frac{p(z|x)}{q^t(z)} \frac{1}{Z}} \right)$$

$$\text{RELBO} \quad \arg \max_{s(z) \in \mathcal{Q}} \underbrace{-\lambda \mathbb{E}_s[\log s(z)] + \mathbb{E}_s[\log p(x, z)]}_{\substack{\text{ELBO} \\ (\text{if } \lambda=1)}} - \underbrace{\mathbb{E}_s[\log q^t(z)]}_{\text{Residual}}$$

- The next component should be a good approximation of the posterior
- But should be different from our current approximation

Results: Proof of Concept



- o Edward implementation: <https://github.com/ratschlab/boosting-bbvi>
- o Easy to try different models and different families
- o Optimization independent from modeling choices

Thanks

Poster: Room 210 & 230 AB #38

Paper: <https://arxiv.org/abs/1806.02185>

Gunnar Rätsch is hiring ML postdocs:
goo.gl/u9UTTo

ETH zürich



MAX-PLANCK-GESellschaft

Max Planck ETH Center
for Learning Systems



ETH zürich

Berkeley
UNIVERSITY OF CALIFORNIA