Proximity Variational Inference

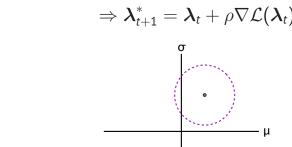
Jaan Altosaar, Rajesh Ranganath, David Blei Princeton University, Columbia University

 $\mathcal{L}(\lambda) = \mathbb{E}_{q(\mathbf{z}; \lambda)}[\log p(\mathbf{x}, \mathbf{z}) - \log q(\mathbf{z}; \lambda)]$

Gradient ascent using proximity operators

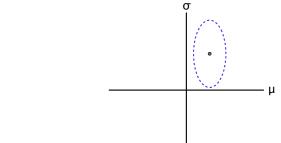
$$egin{aligned} U(oldsymbol{\lambda}_{t+1}) = & \mathcal{L}(oldsymbol{\lambda}_t) +
abla \mathcal{L}(oldsymbol{\lambda}_t)^ op (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t) \end{aligned}$$

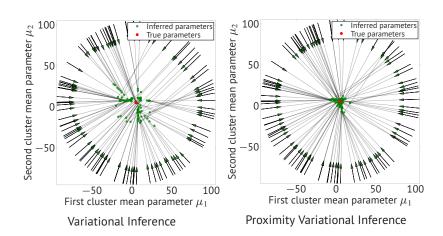
$$egin{aligned} \mathcal{U}(oldsymbol{\lambda}_{t+1}) = & \mathcal{L}(oldsymbol{\lambda}_t) +
abla \mathcal{L}(oldsymbol{\lambda}_t)^{ op} (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t) \ & - rac{1}{2
ho} (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t)^{ op} (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t) \ & \Rightarrow oldsymbol{\lambda}_{t+1}^* = oldsymbol{\lambda}_t +
ho
abla \mathcal{L}(oldsymbol{\lambda}_t) \end{aligned}$$



Proximity operators for variational inference

$$egin{aligned} extstyle U(oldsymbol{\lambda}_{t+1}) = & \mathcal{L}(oldsymbol{\lambda}_t) +
abla \mathcal{L}(oldsymbol{\lambda}_t)^ op (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t) \ & - rac{1}{2
ho} (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t)^ op (oldsymbol{\lambda}_{t+1} - oldsymbol{\lambda}_t) \ & - ext{kd} (f(oldsymbol{\lambda}_t), f(oldsymbol{\lambda}_{t+1})) \end{aligned}$$





Binarized MNIST



(a) Proximity Variational Inference (b) Data (c) Variational Inference