Natural Gradients via the Variational Predictive Distribution

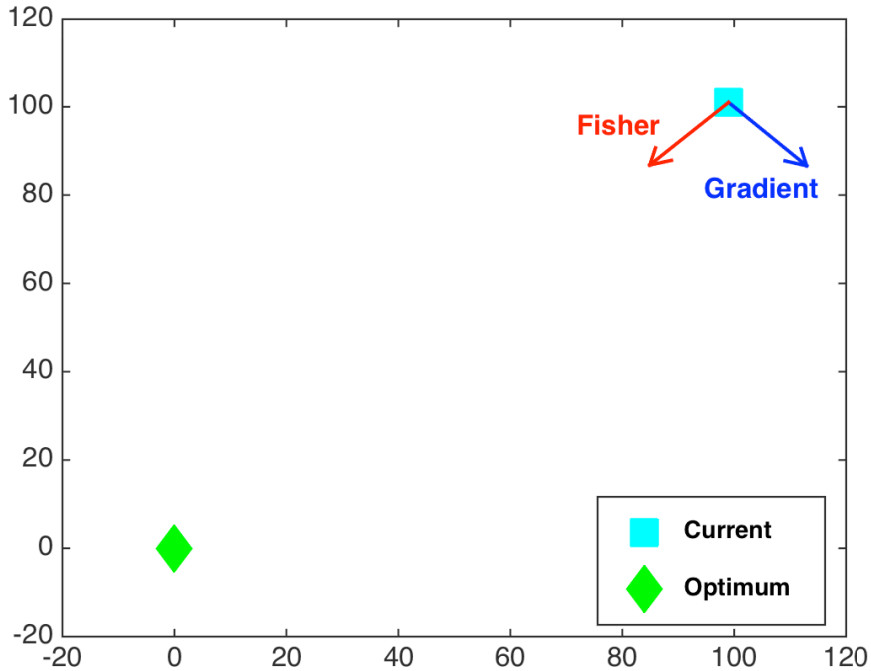
Da Tang[1] and Rajesh Ranganath[2]

Columbia University[1]
New York University[2]

December 8, 2017

Pathological Curvature of the ELBO

- The curvature of the ELBO may be pathological
- Example: A bivariate Gaussian model with unknown mean and known covariance $\Sigma = \begin{pmatrix} 1 & 1-\varepsilon \\ 1-\varepsilon & 1 \end{pmatrix}$, $0 < \varepsilon \ll 1$



- The natural gradient (Hoffman et al., 2013) fails to help

The Variational Predictive Natural Gradient

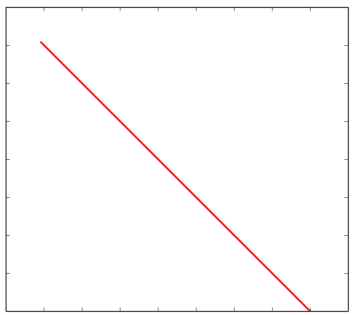
- Approximate the posterior predictive distribution:

$$r(\boldsymbol{x}'|\boldsymbol{x}_i;\boldsymbol{\lambda}) = \int p(\boldsymbol{x}'|\boldsymbol{z}_i,\boldsymbol{\beta})q(\boldsymbol{z}_i|\boldsymbol{x}_i,\boldsymbol{\beta};\boldsymbol{\lambda})q(\boldsymbol{\beta};\boldsymbol{\lambda})d\boldsymbol{z}_i d\boldsymbol{\beta}$$

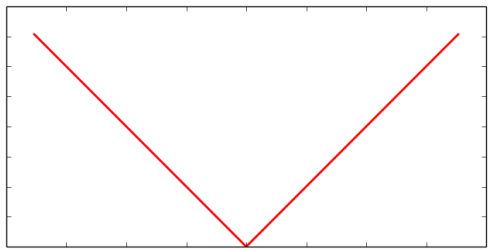
- The variational predictive Fisher information:

$$F_r = \mathbb{E}_{Q_{\boldsymbol{x}_i}, r(\boldsymbol{x}'|\boldsymbol{x}_i;\boldsymbol{\lambda})}[\nabla_{\boldsymbol{\lambda}} \log r(\boldsymbol{x}'|\boldsymbol{x}_i;\boldsymbol{\lambda}) \cdot \nabla_{\boldsymbol{\lambda}} \log r(\boldsymbol{x}'|\boldsymbol{x}_i;\boldsymbol{\lambda})^\top]$$

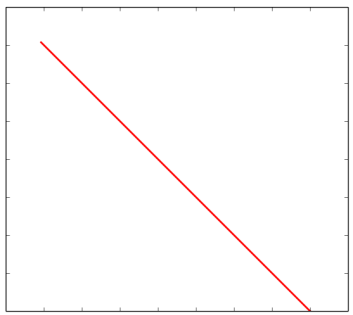
- Eigenspace comparison:



(a) Precision mat Σ^{-1}



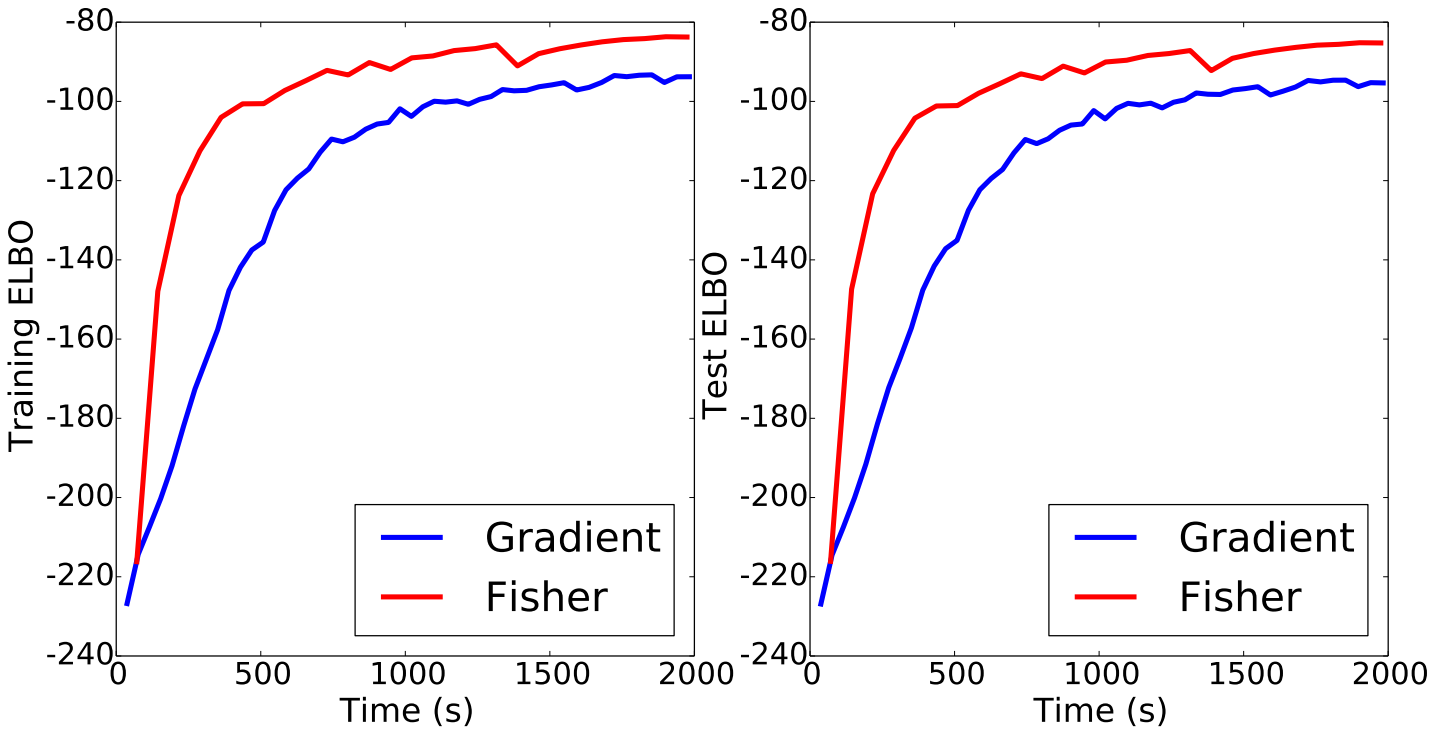
(b) q-Fisher info F_q



(c) Our Fisher info F_r

Experiment: Learning a VAE

- Training on the MNIST dataset (Lecun et al., 1998)
- Learning curves:



Thank you!