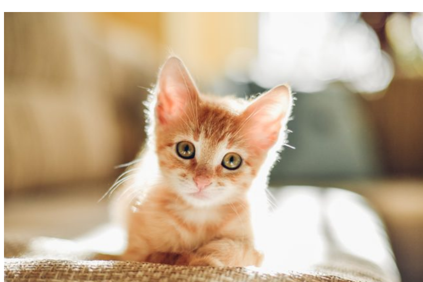
AUTO-ENCODING VARIATIONAL BAYES – KINGMA, WELLING 2014

Darshan Thaker

October 3, 2017

Problem Statement

Input: **X**



Model P(X)

Problem Statement

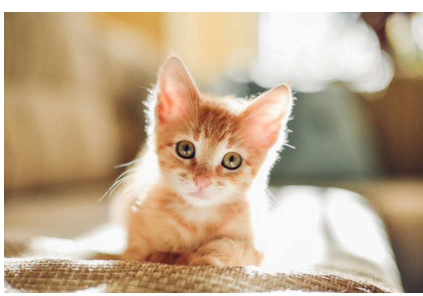
Input: **X**

Model P(X)

Latent Variables z

Problem Statement

Input: **X**



Model P(Z | X)

Latent Variables z

Why is this hard?

Assume from some parametric family

$$P(z|x) = \frac{P(x|z)P(z)}{P(x)}$$

Typically intractable integral

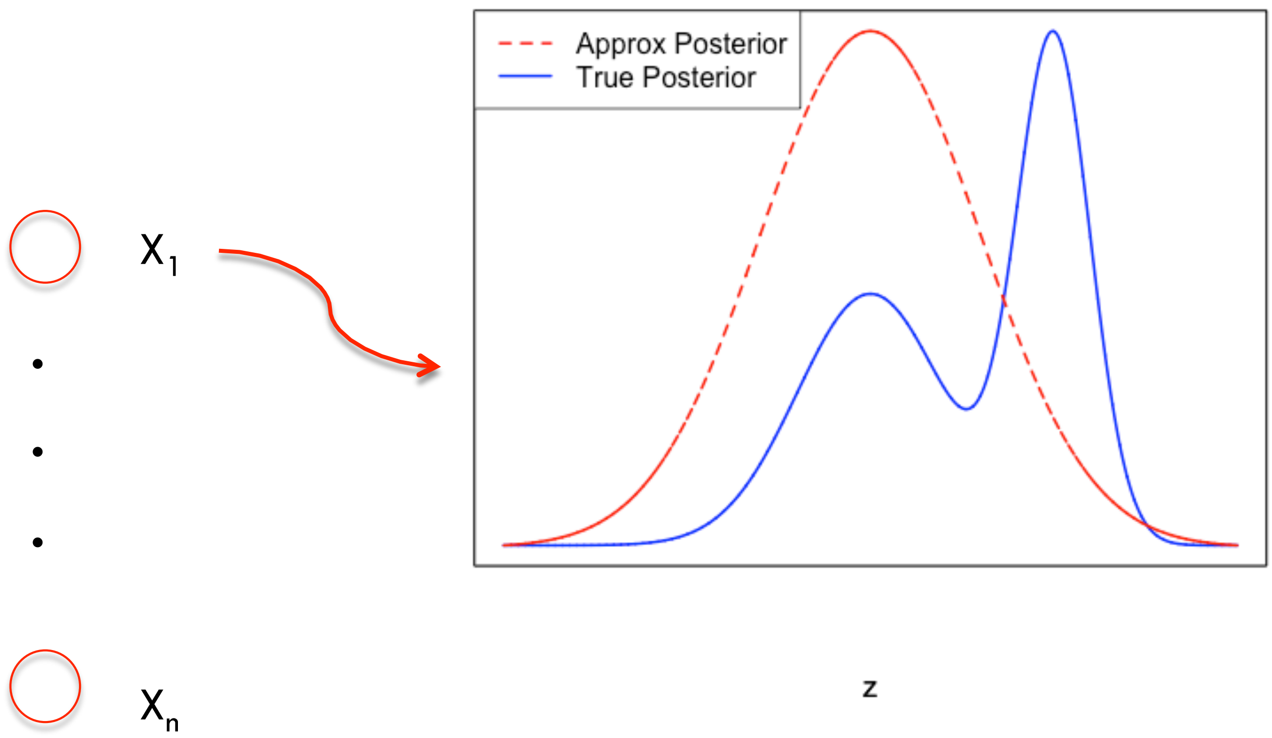
Why is this hard?

$$P(z|x) = \frac{P(x|z)P(z)}{P(x)}$$

$$q_\phi(z|x) \sim P(z|x)$$

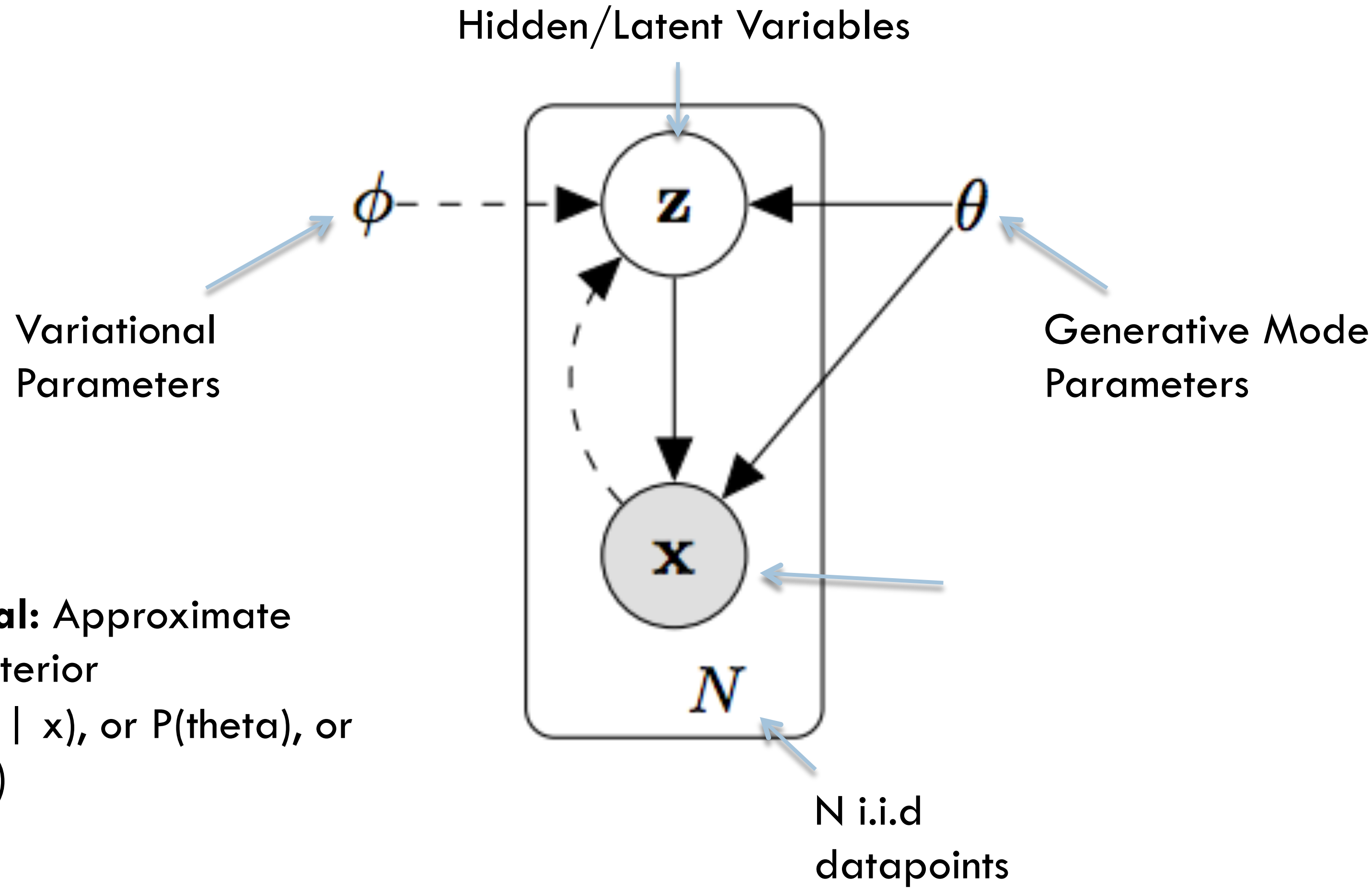
Variational parameters

More background



Goal: Minimize KL(approx, true)

Generative model



Goal: Approximate posterior
P(z | x), or P(theta), or P(x)

Variational Lower Bound

$$\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) \geq \boxed{\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}^{(i)})} = \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}\left[-\log q_{\phi}(\mathbf{z}|\mathbf{x}) + \log p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z})\right]$$

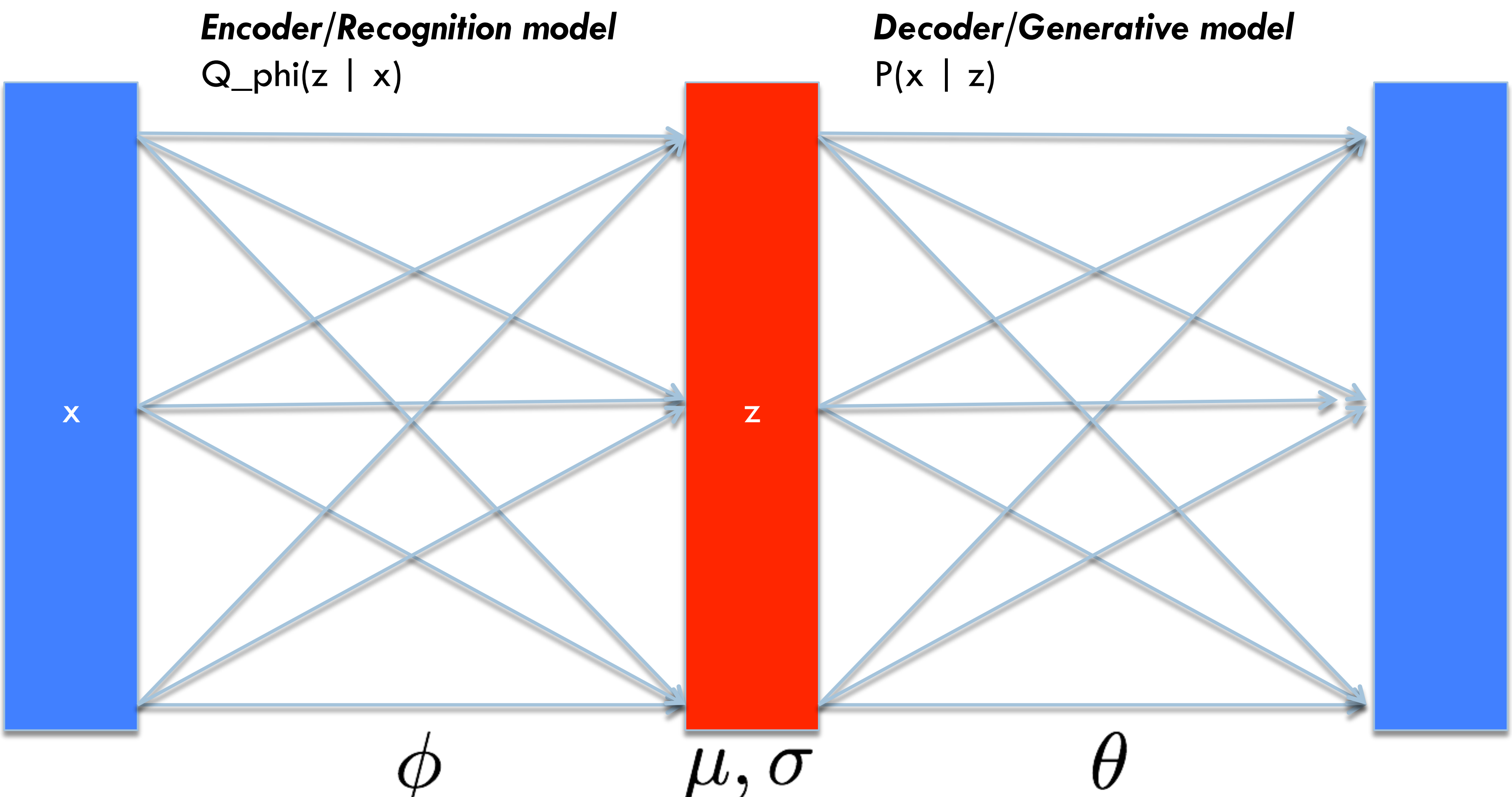
$$= -D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)}) || p_{\boldsymbol{\theta}}(\mathbf{z})) + \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\right]$$

Log-Likelihood of datapoint

Penalize deviation of approximation from prior

Reward fit to data

Bring in the Autoencoder!



Backpropagation through sampling

$Z^1 \sim q(z \mid x)$

.

.

.

$Z^n \sim q(z \mid x)$

Monte Carlo Gradient

$$\frac{1}{L}\sum_{l=1}^{L} f(\mathbf{z})\nabla_{q_\phi(\mathbf{z}^{(l)})}\log q_\phi(\mathbf{z}^{(l)})$$

Backpropagation through sampling

$Z^1 \sim q(z \mid x)$

.

.

.

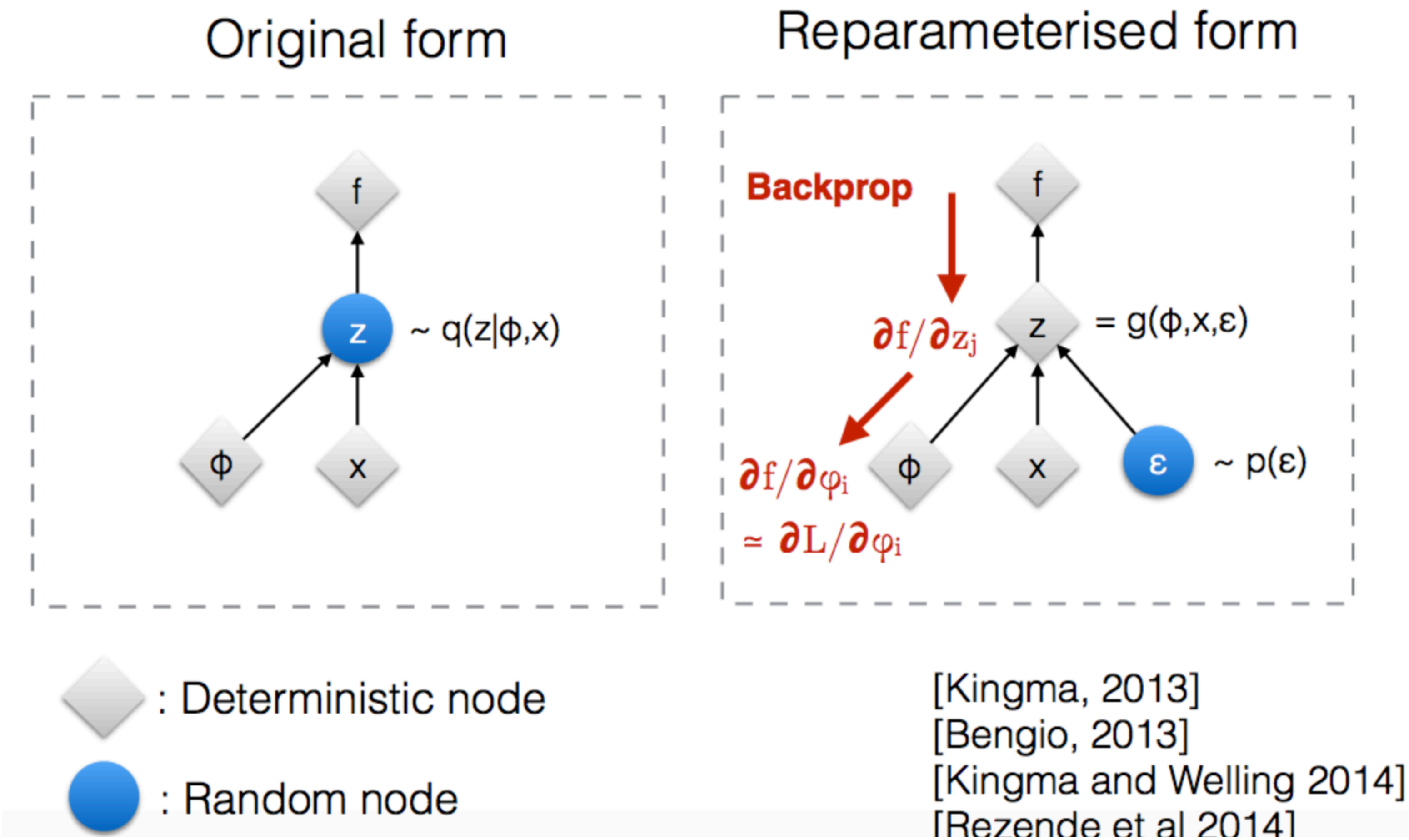
$Z^n \sim q(z \mid x)$

Monte Carlo Gradient

$$\frac{1}{L}\sum_{l=1}^{L} f(\mathbf{z})\nabla_{q_\phi(\mathbf{z}^{(l)})} \log q_\phi(\mathbf{z}^{(l)})$$

High Variance estimate!

Reparametrization Trick + SGVB



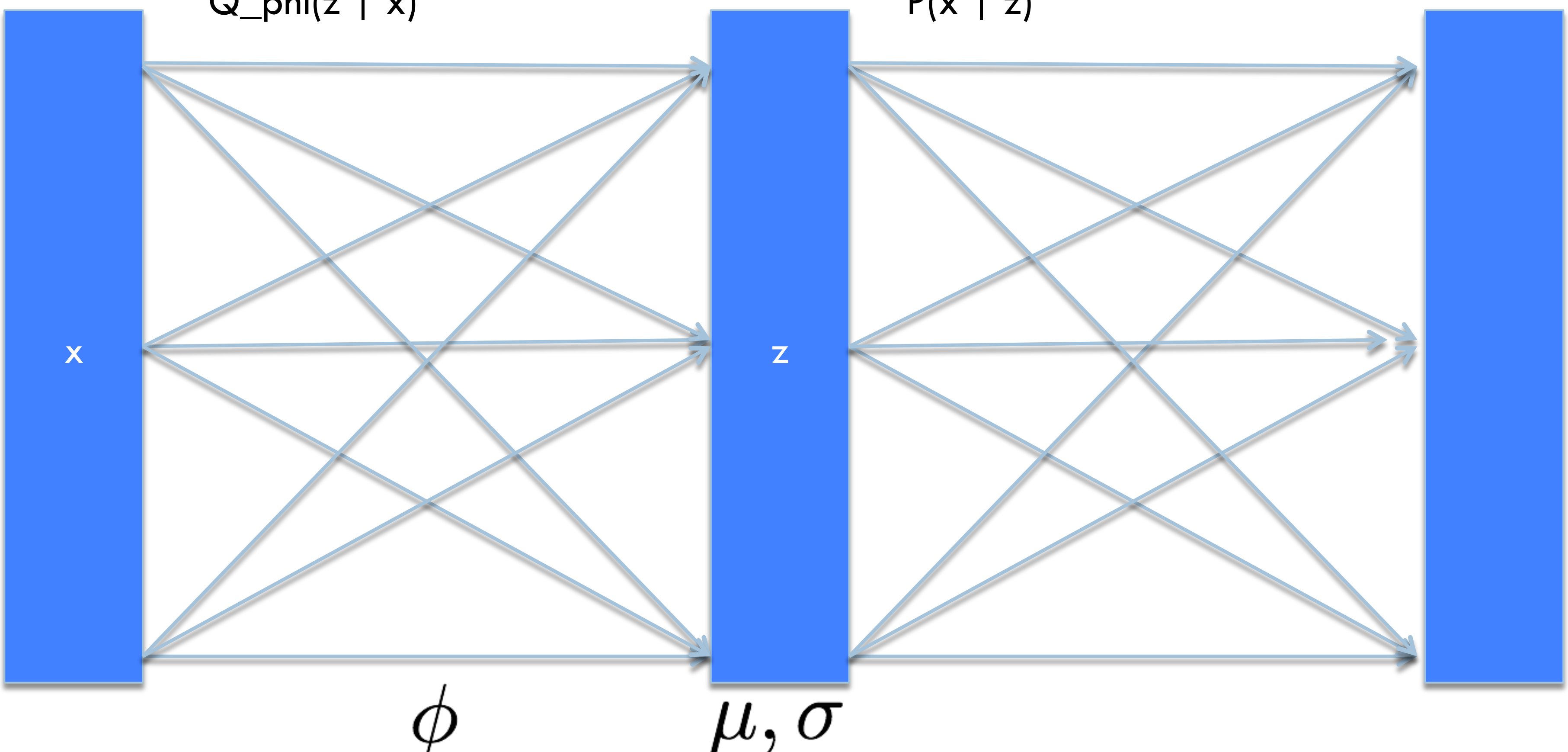
Slide credit: Kingma NIPS 2015 workshop Slides

Variational Auto-encoders

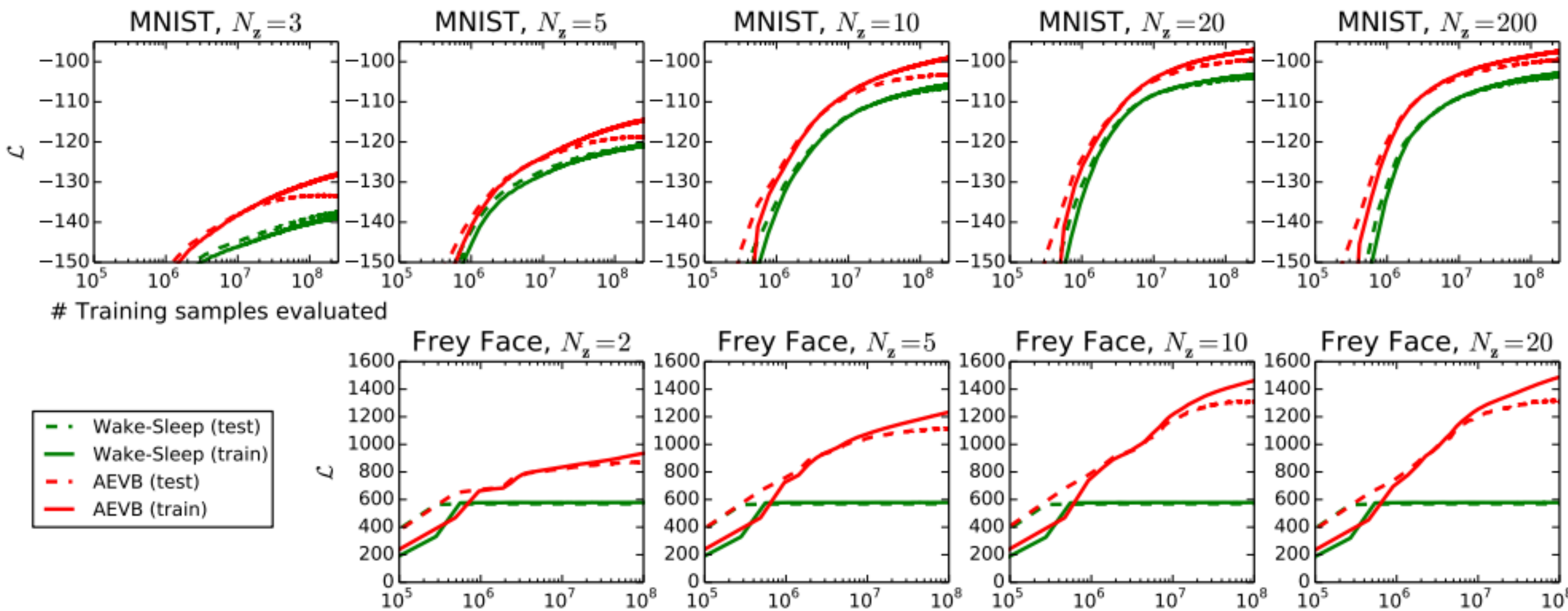
Optimize with stochastic gradient ascent using SGVB estimator + gradients

Encoder/Recognition model
Q_phi(z | x)

Decoder/Generative model
P(x | z)



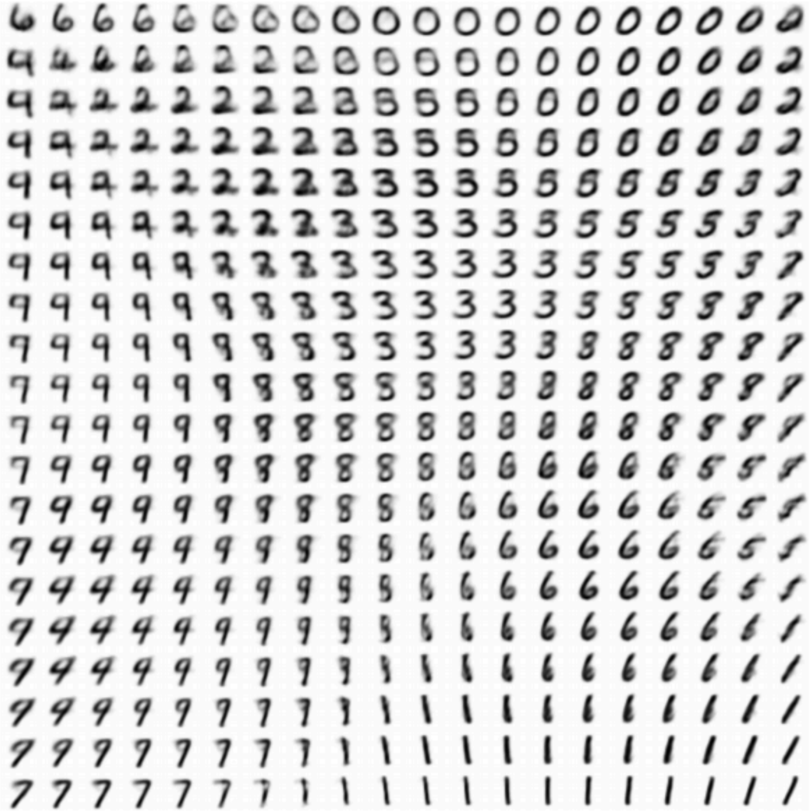
Results



MNIST, $N_z = 3$
MNIST, $N_z = 5$
MNIST, $N_z = 10$
MNIST, $N_z = 20$
MNIST, $N_z = 200$
$\mathcal{L}$
−100
−110
−120
−130
−140
−150
10^5
10^6
10^7
10^8
Training samples evaluated
Frey Face, $N_z = 2$
Frey Face, $N_z = 5$
Frey Face, $N_z = 10$
Frey Face, $N_z = 20$
1600
1400
1200
1000
800
600
400
200
0
Wake-Sleep (test)
Wake-Sleep (train)
AEVB (test)
AEVB (train)

Results

(a) Learned Frey Face manifold



(b) Learned MNIST manifold

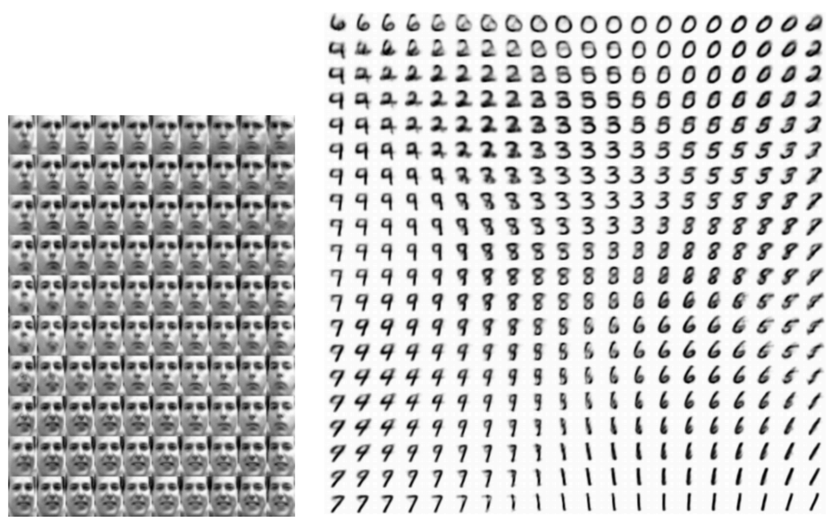
Results

(a) 2-D latent space (b) 5-D latent space (c) 10-D latent space (d) 20-D latent space

Pros

Fascinating idea

- Connection between auto-encoders and variational inference
- Good results in generating real-world datasets (faces and MNIST)



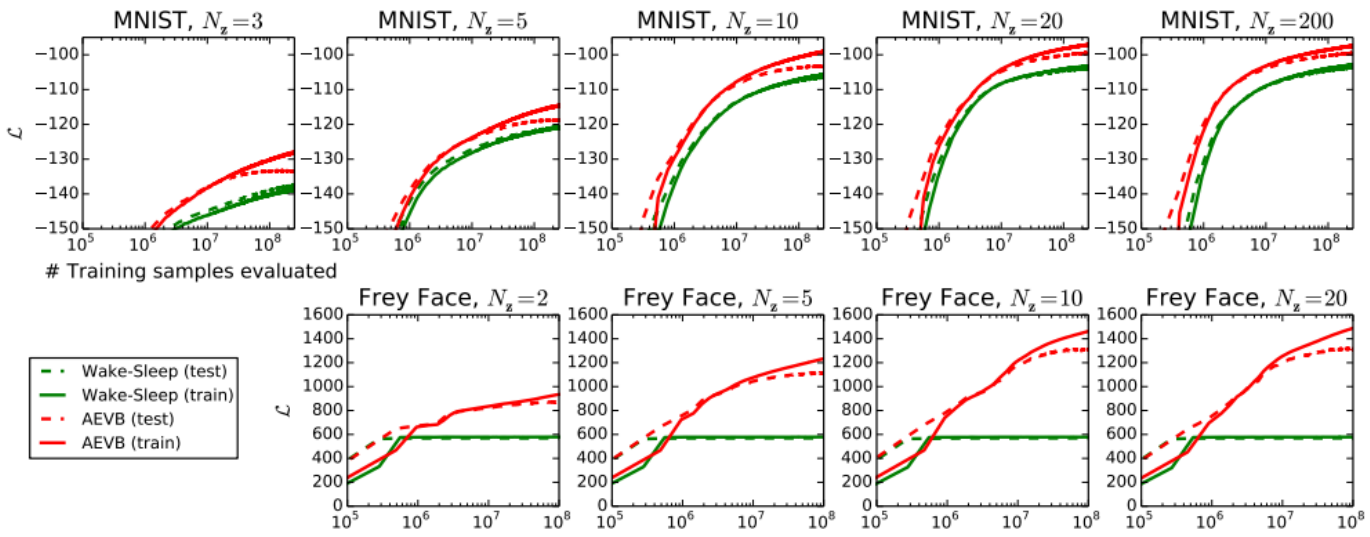
(a) Learned Frey Face manifold (b) Learned MNIST manifold

Practical

- Proposes real solution to problem of intractable integrals while computing posterior
- Classic Bayesian model

Overfitting

- Regularizing nature of lower bound



Latent Variable Space