

# Implicit Reparameterization Gradients

**Michael Figurnov, Shakir Mohamed, Andriy Mnih**

**Poster: Room 210 #33**



**DeepMind**

# Reparameterization gradients

Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

# Reparameterization gradients

Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

$$\frac{\partial}{\partial \phi} \mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)} \left[ \frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial \phi} \right]$$

continuous  
(Normal, ...)

differentiable  
(ELBO, ...)

backpropagation

# Reparameterization gradients

Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

$$\frac{\partial}{\partial \phi} \mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)} \left[ \frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial \phi} \right]$$

continuous  
(Normal, ...)      differentiable  
(ELBO, ...)      backpropagation

requires a *tractable* inverse transformation!  
Normal, Logistic, ...

A mathematical equation illustrating the backpropagation rule for reparameterization gradients. The left side shows the derivative of the expectation of a function f(z) with respect to parameters phi. The right side is an expectation over the latent variable z, multiplied by a Jacobian determinant. This determinant is composed of two ratios: the derivative of the function f(z) with respect to z, and the derivative of z with respect to phi. The second ratio is highlighted with a red box. Below the equation, two arrows point upwards from the text 'continuous (Normal, ...)' and 'differentiable (ELBO, ...)' to the first and second terms of the expectation respectively. To the right of the equation, a callout box contains the text 'requires a *tractable* inverse transformation!' followed by examples like 'Normal, Logistic, ...'.

# Reparameterization gradients

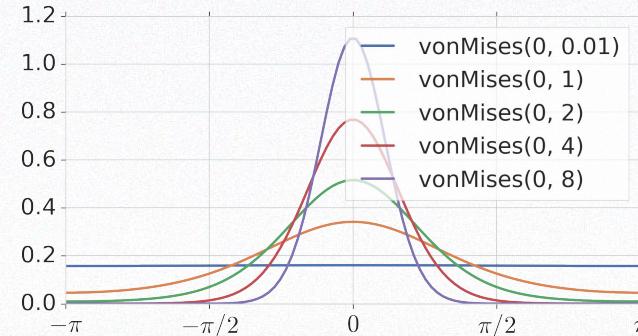
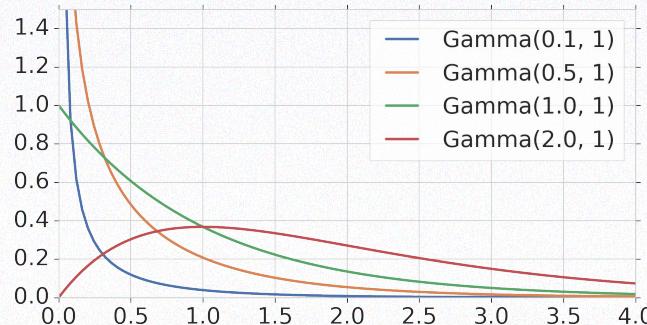
Core part of variational autoencoders, automatic variational inference, etc.

Backpropagation in graphs with continuous random variables

$$\frac{\partial}{\partial \phi} \mathbb{E}_{q_\phi(z)}[f(z)] = \mathbb{E}_{q_\phi(z)} \left[ \frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial \phi} \right]$$

requires a *tractable* inverse transformation!  
Normal, Logistic, ...

We show how to use *implicit differentiation* for reparameterization of other continuous random variables, such as Gamma and von Mises



# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

Sampling (forward pass)

Gradients (backward pass)

---

Explicit

---

Implicit

# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

Sampling (forward pass)

Gradients (backward pass)

---

Explicit

$u \sim \text{Uniform}(0, 1)$

$z = F^{-1}(u|\phi) \sim q_\phi(z)$

---

Implicit

# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0, 1)$ $z = F^{-1}(u \phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u \phi)}{\partial \phi}$
Implicit		

# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0, 1)$ $z = F^{-1}(u \phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u \phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	

# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0, 1)$ $z = F^{-1}(u \phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u \phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	$\frac{\partial z}{\partial \phi} = -\frac{\frac{\partial F(z \phi)}{\partial \phi}}{q_\phi(z)}$

# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0, 1)$ $z = F^{-1}(u \phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u \phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	$\frac{\partial z}{\partial \phi} = -\frac{\frac{\partial F(z \phi)}{\partial \phi}}{q_\phi(z)}$ <p>Derivation: implicit differentiation <math>\frac{dF(z \phi)}{d\phi} = \frac{du}{d\phi}</math></p> $q_\phi(z) \xrightarrow{\frac{\partial F(z \phi)}{\partial z}} \frac{\partial z}{\partial \phi} + \frac{\partial F(z \phi)}{\partial \phi} = 0$

# Explicit and implicit reparameterization

Cumulative density function  $F(z|\phi) \equiv \int_{-\infty}^z q_\phi(t)dt = u \sim \text{Uniform}(0, 1)$

	Sampling (forward pass)	Gradients (backward pass)
Explicit	$u \sim \text{Uniform}(0, 1)$ $z = F^{-1}(u \phi) \sim q_\phi(z)$	$\frac{\partial z}{\partial \phi} = \frac{\partial F^{-1}(u \phi)}{\partial \phi}$
Implicit	$z \sim q_\phi(z)$ using any sampler (e.g., rejection sampling)	$\frac{\partial z}{\partial \phi} = -\frac{\frac{\partial F(z \phi)}{\partial \phi}}{q_\phi(z)}$ <p>Derivation: implicit differentiation <math>\frac{dF(z \phi)}{d\phi} = \frac{du}{d\phi}</math></p> $q_\phi(z) \frac{\partial F(z \phi)}{\partial z} \frac{\partial z}{\partial \phi} + \frac{\partial F(z \phi)}{\partial \phi} = 0$ <p>often not implemented in numerical libraries</p>

# How to compute $\frac{\partial F(z|\phi)}{\partial \phi}$ ?

Relative metrics (lower is better)

Method	Gamma		Von Mises	
	Error	Time	Error	Time
<b>Automatic differentiation of the CDF code</b>	<b>1x</b>	<b>1x</b>	<b>1x</b>	<b>1x</b>
Finite difference	832x	2x	514x	1.2x
Jankowiak & Obermeyer (2018) concurrent work; closed-form approximation	18x	5x	-	-

Jankowiak, Obermeyer "Pathwise Derivatives Beyond the Reparameterization Trick." ICML, 2018

# How to compute $\frac{\partial F(z|\phi)}{\partial \phi}$ ?

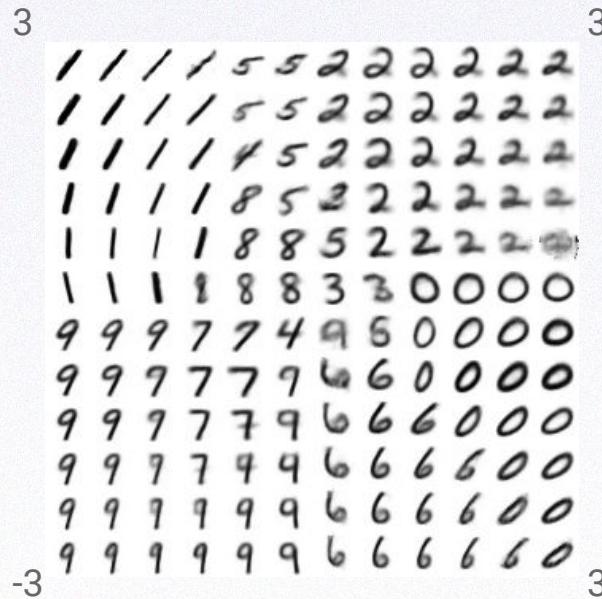
Relative metrics (lower is better)

Method	Gamma		Von Mises	
	Error	Time	Error	Time
<b>Automatic differentiation of the CDF code</b>	1x	1x	1x	1x
Finite difference	832x	2x	514x	1.2x
Jankowiak & Obermeyer (2018) concurrent work; closed-form approximation	18x	5x	-	-
Knowles (2015) approximate <i>explicit</i> reparameterization	2840x	63x	-	-

Knowles, "Stochastic gradient variational Bayes for Gamma approximating distributions." arXiv, 2015  
Jankowiak, Obermeyer "Pathwise Derivatives Beyond the Reparameterization Trick." ICML, 2018

# Variational Autoencoder

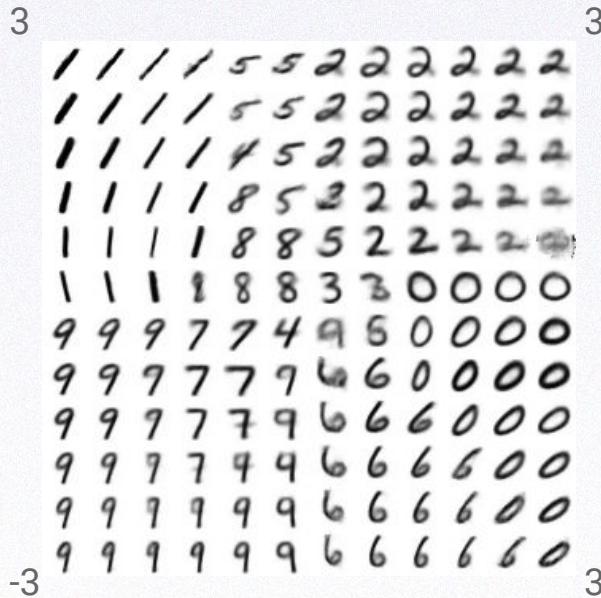
2D latent spaces for MNIST



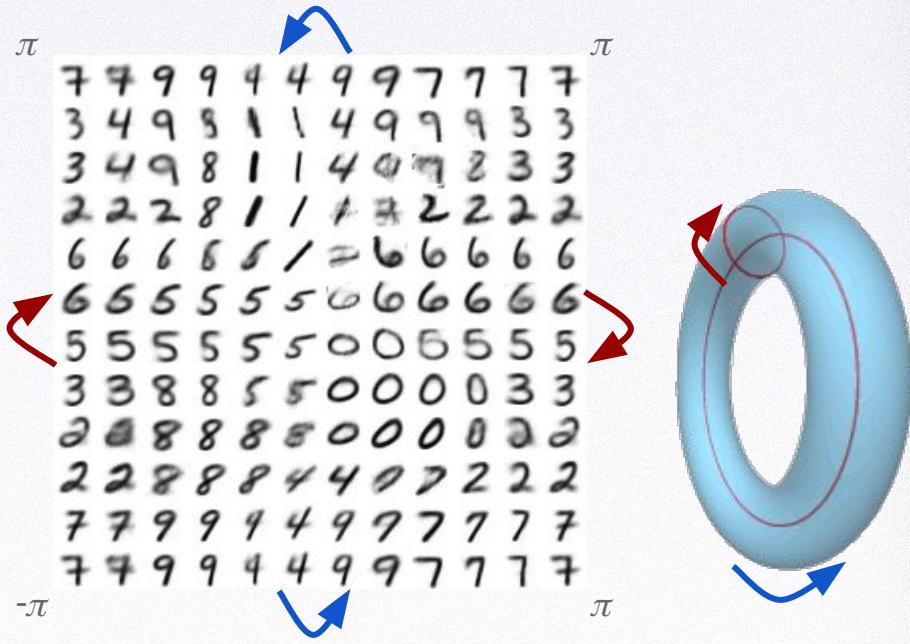
Normal prior and posterior

# Variational Autoencoder

2D latent spaces for MNIST



Normal prior and posterior



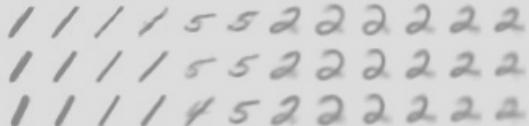
Uniform prior, von Mises posterior

Torus adapted from [https://en.wikipedia.org/wiki/Torus#/media/File:Sphere-like\\_degenerate\\_torus.gif](https://en.wikipedia.org/wiki/Torus#/media/File:Sphere-like_degenerate_torus.gif)

# Variational Autoencoder

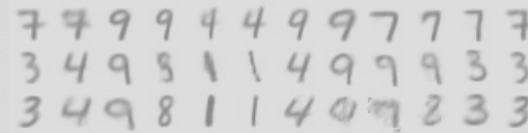
2D latent spaces for MNIST

3



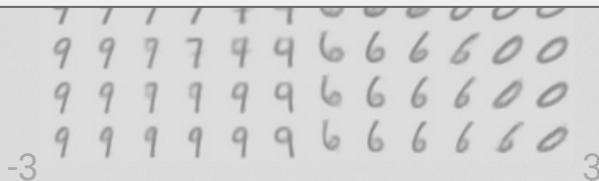
3

$\pi$



$\pi$

Also in the paper: **Latent Dirichlet Allocation**



-3

Normal prior and posterior



$-\pi$

Uniform prior, von Mises posterior

Torus adapted from [https://en.wikipedia.org/wiki/Torus#/media/File:Sphere-like\\_degenerate\\_torus.gif](https://en.wikipedia.org/wiki/Torus#/media/File:Sphere-like_degenerate_torus.gif)

# Implicit Reparameterization Gradients

Michael Figurnov, Shakir Mohamed, Andriy Mnih

- A more general view of the reparameterization gradients
  - Decouple sampling from gradient estimation
- Reparameterization gradients for Gamma, von Mises, Beta, Dirichlet, ...
  - Faster and more accurate than the alternatives
  - Implemented in TensorFlow Probability:  
`tfp.distributions.{Gamma, VonMises, Beta, Dirichlet, ...}`
- Move away from making modelling choices for computational convenience

Poster: Room 210 #33