Continuous Normalizing Flows

Previously on Seminar

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) - initial\ state$$

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) - initial\ state$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) - initial\ state$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$

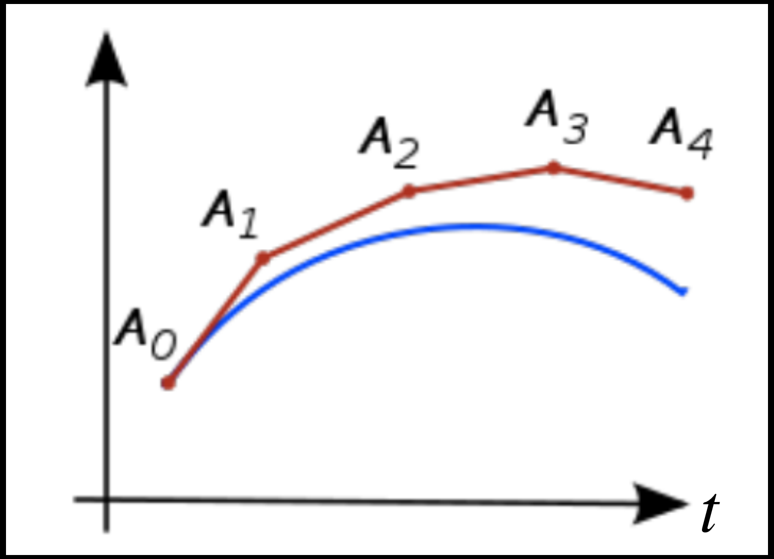
Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) - initial\ state$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



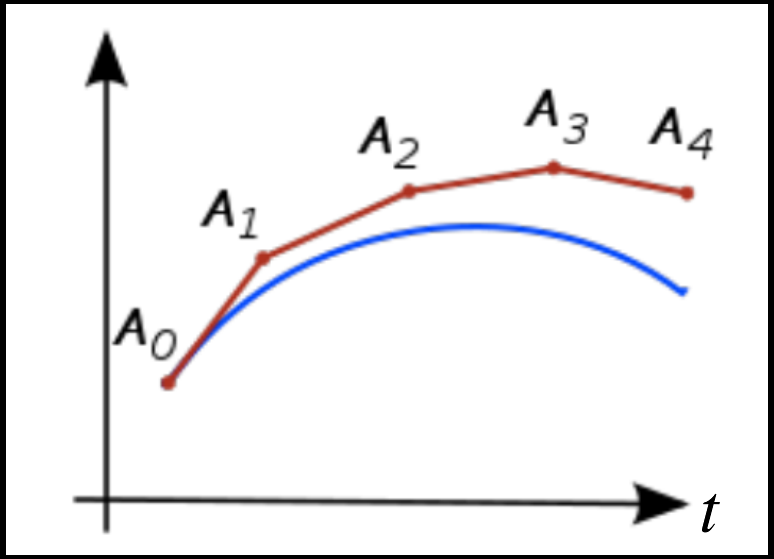
Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$\vec{z}(0) = input$ $\qquad$ $\vec{z}(T) = ouput$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Neural ODE

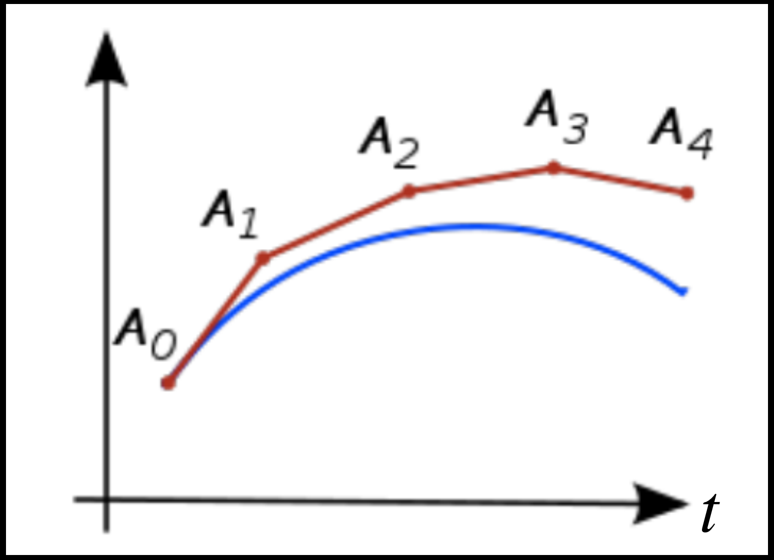
$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \; - \; loss$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

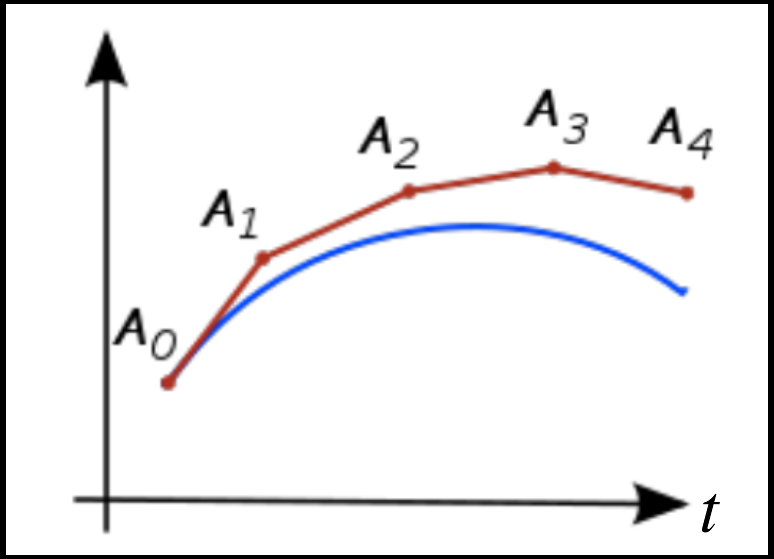
$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \ - \ loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \overrightarrow{\theta}} = ?$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t) dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \ - \ loss$$

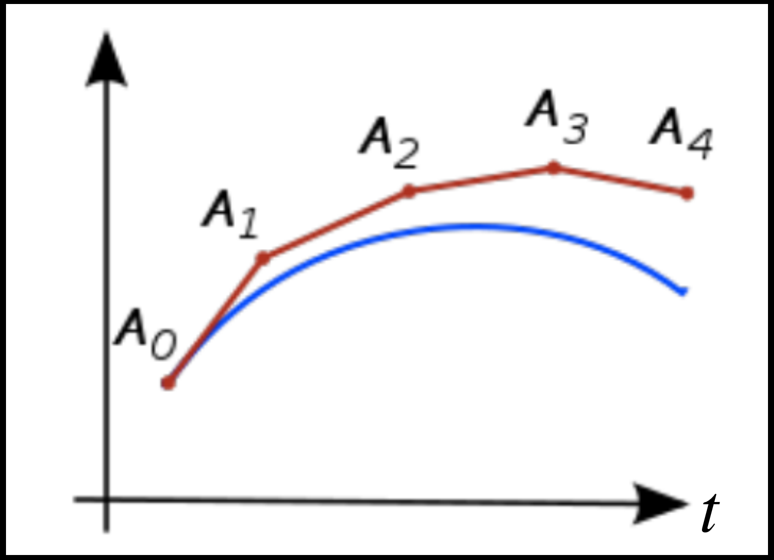
$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \overrightarrow{\theta}} = ?$$

Adjoint system

$$\overrightarrow{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \ - \ loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \overrightarrow{\theta}} = ?$$

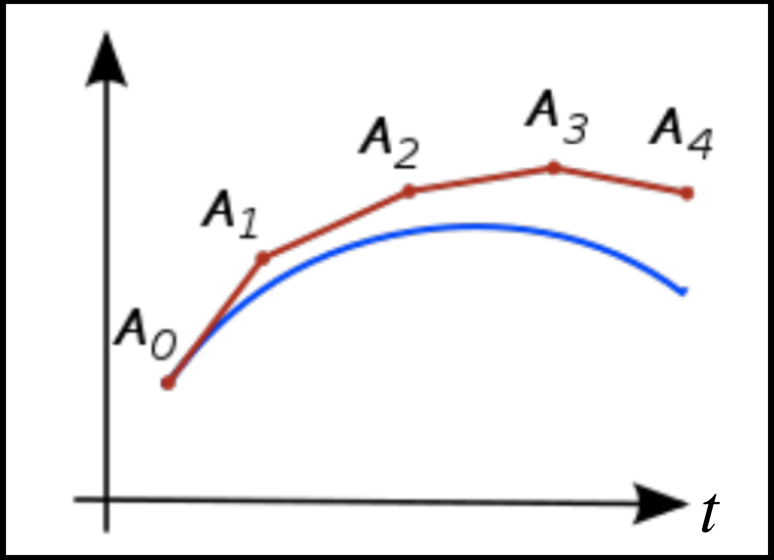
Adjoint system

$$\overrightarrow{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\overrightarrow{a}(t)}{dt} = - \left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \vec{z}(t)}\right]^T \overrightarrow{a}(t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \ - \ loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \overrightarrow{\theta}} = ?$$

Adjoint system

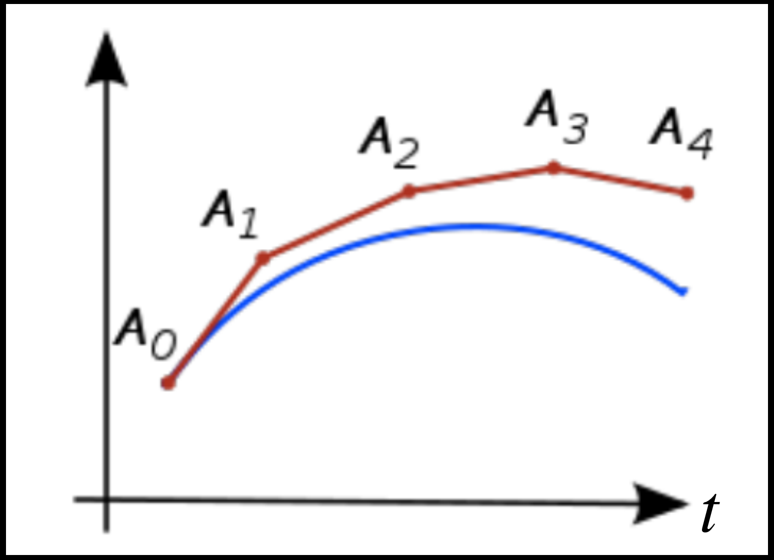
$$\overrightarrow{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\overrightarrow{a}(t)}{dt} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \vec{z}(t)}\right]^T \overrightarrow{a}(t)$$

$$\frac{d}{dt}\frac{\partial L}{\partial \overrightarrow{\theta}(t)} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \overrightarrow{\theta}}\right]^T \overrightarrow{a}(t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \; - \; loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \overrightarrow{\theta}} = ?$$

Adjoint system

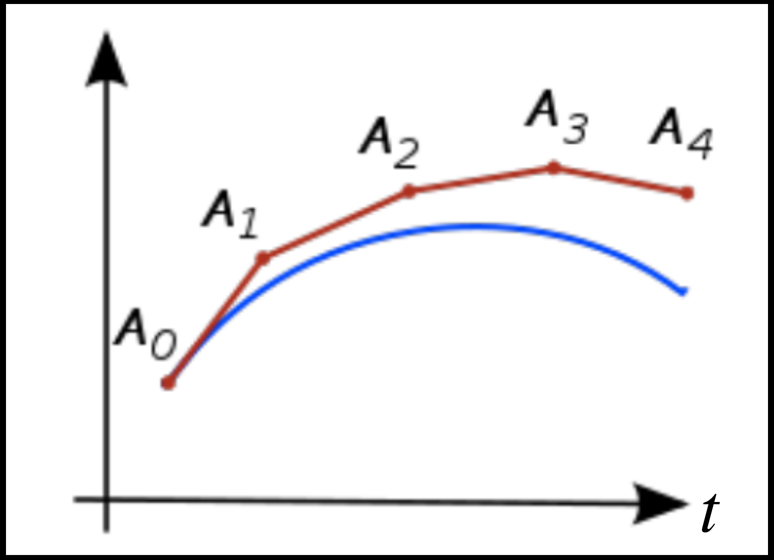
$$\overrightarrow{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\overrightarrow{a}(t)}{dt} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \vec{z}(t)}\right]^T \overrightarrow{a}(t)$$

$$\frac{d}{dt}\frac{\partial L}{\partial \overrightarrow{\theta}(t)} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \overrightarrow{\theta}}\right]^T \overrightarrow{a}(t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Training

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \vec{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \; - \; loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \vec{\theta}} = ?$$

Adjoint system

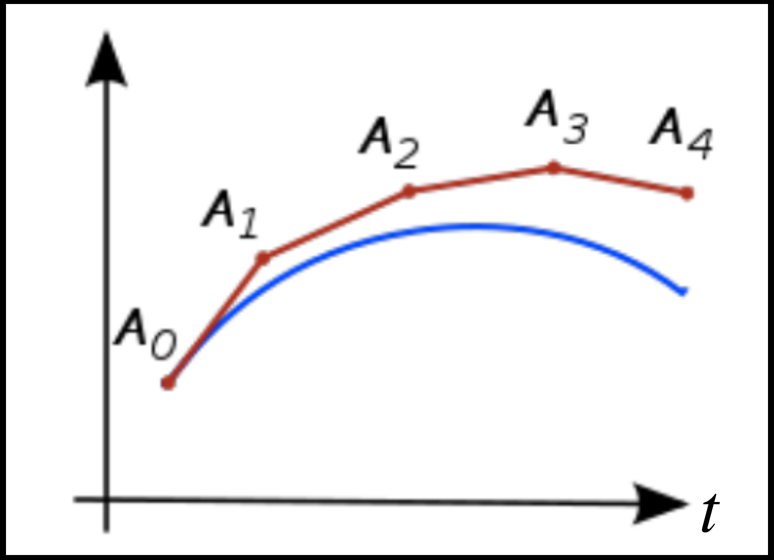
$$\vec{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\vec{a}(t)}{dt} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{z}(t)}\right]^T \vec{a}(t)$$

$$\frac{d}{dt}\frac{\partial L}{\partial \vec{\theta}(t)} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{\theta}}\right]^T \vec{a}(t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \vec{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \vec{\theta})$$



Training

$$\vec{z}(0)$$

Neural ODE

$$\boxed{\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \ - \ loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \overrightarrow{\theta}} = ?$$

Adjoint system

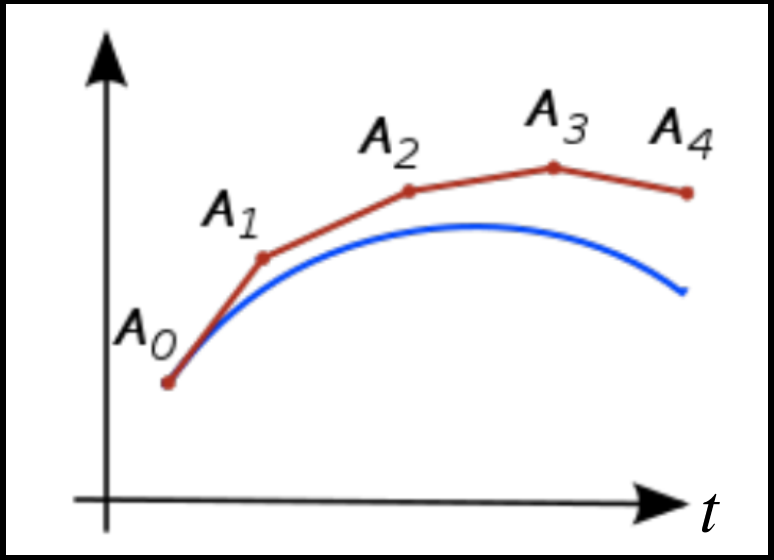
$$\overrightarrow{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\overrightarrow{a}(t)}{dt} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \vec{z}(t)}\right]^T \overrightarrow{a}(t)$$

$$\frac{d}{dt}\frac{\partial L}{\partial \overrightarrow{\theta}(t)} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)}{\partial \overrightarrow{\theta}}\right]^T \overrightarrow{a}(t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \overrightarrow{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \overrightarrow{\theta})$$



Training

$$\vec{z}(0) \xrightarrow{\text{ODESolver}} \vec{z}(T)$$

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \vec{\theta}, t)$$

$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$\boxed{L(\vec{z}(T)) \; - \; loss}$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \vec{\theta}} = ?$$

Adjoint system

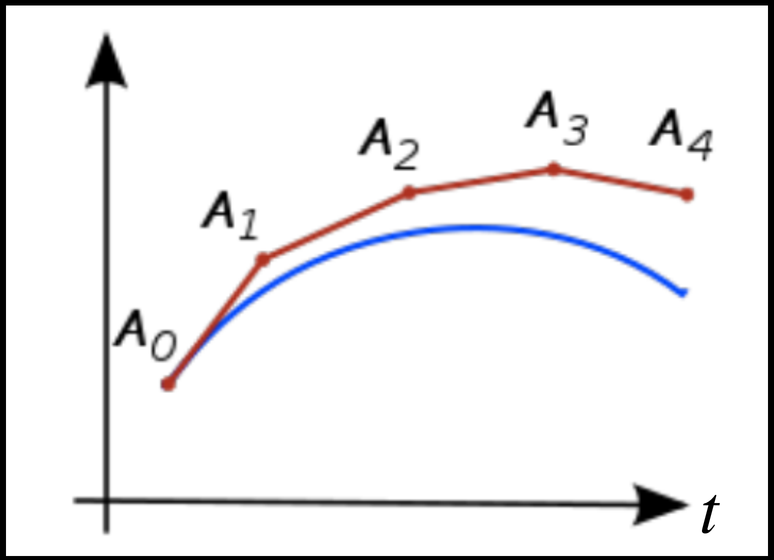
$$\vec{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\vec{a}(t)}{dt} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{z}(t)}\right]^T \vec{a}(t)$$

$$\frac{d}{dt}\frac{\partial L}{\partial \vec{\theta}(t)} = -\left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{\theta}}\right]^T \vec{a}(t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \vec{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \vec{\theta})$$



Training

$$\vec{z}(0) \xrightarrow{\text{ODESolver}} \vec{z}(T)$$

$$\frac{\partial L}{\partial \vec{z}(T)}$$

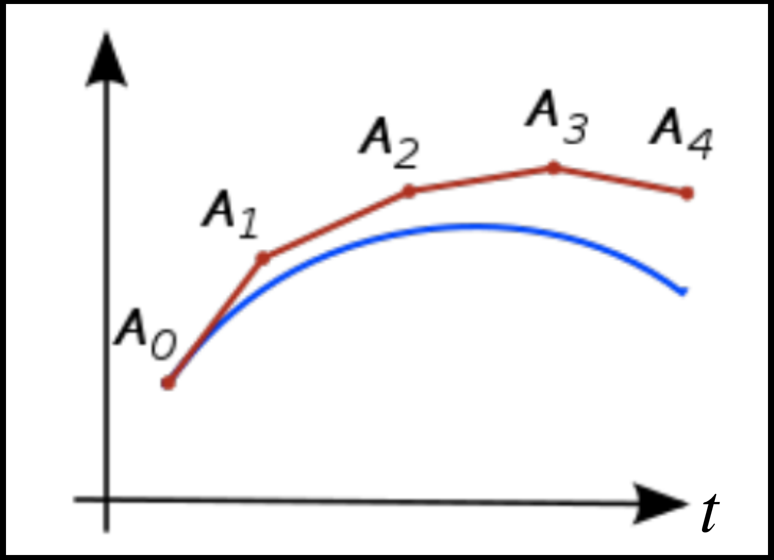
Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \vec{\theta}, t)$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \vec{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \vec{\theta})$$

$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$



$$L(\vec{z}(T)) \; - \; loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \vec{\theta}} = ?$$

Adjoint system

$$\vec{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\frac{d\vec{a}(t)}{dt} = - \left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{z}(t)}\right]^T \vec{a}(t)$$

$$\frac{d}{dt}\frac{\partial L}{\partial \vec{\theta}(t)} = - \left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{\theta}}\right]^T \vec{a}(t)$$

Training

$$\vec{z}(0) \xrightarrow{\text{ODESolver}} \vec{z}(T)$$

$$\frac{\partial L}{\partial \vec{z}(T)}$$

$$\frac{\partial L}{\partial \vec{\theta}(T)} = 0$$

Neural ODE

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), \vec{\theta}, t)$$

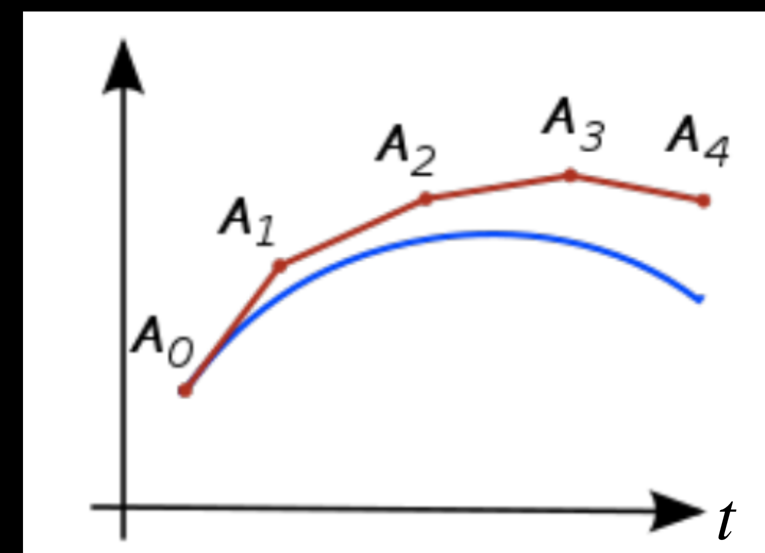
$$\vec{z}(0) = input \qquad \vec{z}(T) = ouput$$

$$L(\vec{z}(T)) \; - \; loss$$

$$\frac{\partial L}{\partial \vec{z}(0)} = ? \qquad \frac{\partial L}{\partial \vec{\theta}} = ?$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), \vec{\theta}, t)dt$$

$$\vec{z}(T) \simeq ODESolver(\vec{z}(0), T, \vec{f}, \vec{\theta})$$



Adjoint system

$$\vec{a}(t) = \frac{\partial L}{\partial \vec{z}(t)}$$

$$\boxed{\begin{aligned} \frac{d\vec{a}(t)}{dt} &= -\left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{z}(t)}\right]^T \vec{a}(t) \\ \frac{d}{dt}\frac{\partial L}{\partial \vec{\theta}(t)} &= -\left[\frac{\partial \vec{f}(\vec{z}(t), \vec{\theta}, t)}{\partial \vec{\theta}}\right]^T \vec{a}(t) \end{aligned}}$$

Training

$$\vec{z}(0) \xrightarrow{ODESolver} \vec{z}(T)$$

$$\left.\begin{matrix} \dfrac{\partial L}{\partial \vec{z}(0)} \\ \dfrac{\partial L}{\partial \vec{\theta}(0)} \end{matrix}\right\} \xleftarrow{ODESolver} \left\{\begin{matrix} \dfrac{\partial L}{\partial \vec{z}(T)} \\ \dfrac{\partial L}{\partial \vec{\theta}(T)} = 0 \end{matrix}\right.$$

Normalizing flow

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0) \qquad \int p_0(\vec{z}_0) d\vec{z}_0 = 1$$

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\int p_0(\vec{z}_0) d\vec{z}_0 = 1$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ is smooth and invertible

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0) \qquad \int p_0(\vec{z}_0) d\vec{z}_0 = 1$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ is smooth and invertible

$$\int p_0(\vec{z}_0) d\vec{z}_0 = \int p_0(\vec{z}_0) \left| \det \frac{\partial \vec{f}^{-1}}{\partial \vec{z}_0} \right| d\vec{z}_1$$

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0) \qquad \int p_0(\vec{z}_0) d\vec{z}_0 = 1$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ is smooth and invertible

$$\int p_0(\vec{z}_0) d\vec{z}_0 = \int p_0(\vec{z}_0) \left| \det \frac{\partial \vec{f}^{-1}}{\partial \vec{z}_0} \right| d\vec{z}_1 = \int p_0(\vec{z}_0) \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|^{-1} d\vec{z}_1 = 1$$

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0) \qquad \int p_0(\vec{z}_0) d\vec{z}_0 = 1$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ is smooth and invertible

$$\int p_0(\vec{z}_0) d\vec{z}_0 = \int p_0(\vec{z}_0) \left| \det \frac{\partial \vec{f}^{-1}}{\partial \vec{z}_0} \right| d\vec{z}_1 = \int \underbrace{p_0(\vec{z}_0) \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|^{-1}}_{p_1(\vec{z}_1)} d\vec{z}_1 = 1$$

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0) \qquad \int p_0(\vec{z}_0)d\vec{z}_0 = 1$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ is smooth and invertible

$$\int p_0(\vec{z}_0)d\vec{z}_0 = \int p_0(\vec{z}_0)\left|\det\frac{\partial \vec{f}^{-1}}{\partial \vec{z}_0}\right| d\vec{z}_1 = \int \underbrace{p_0(\vec{z}_0)\left|\det\frac{\partial \vec{f}}{\partial \vec{z}_0}\right|^{-1}}_{p_1(\vec{z}_1)} d\vec{z}_1 = 1$$

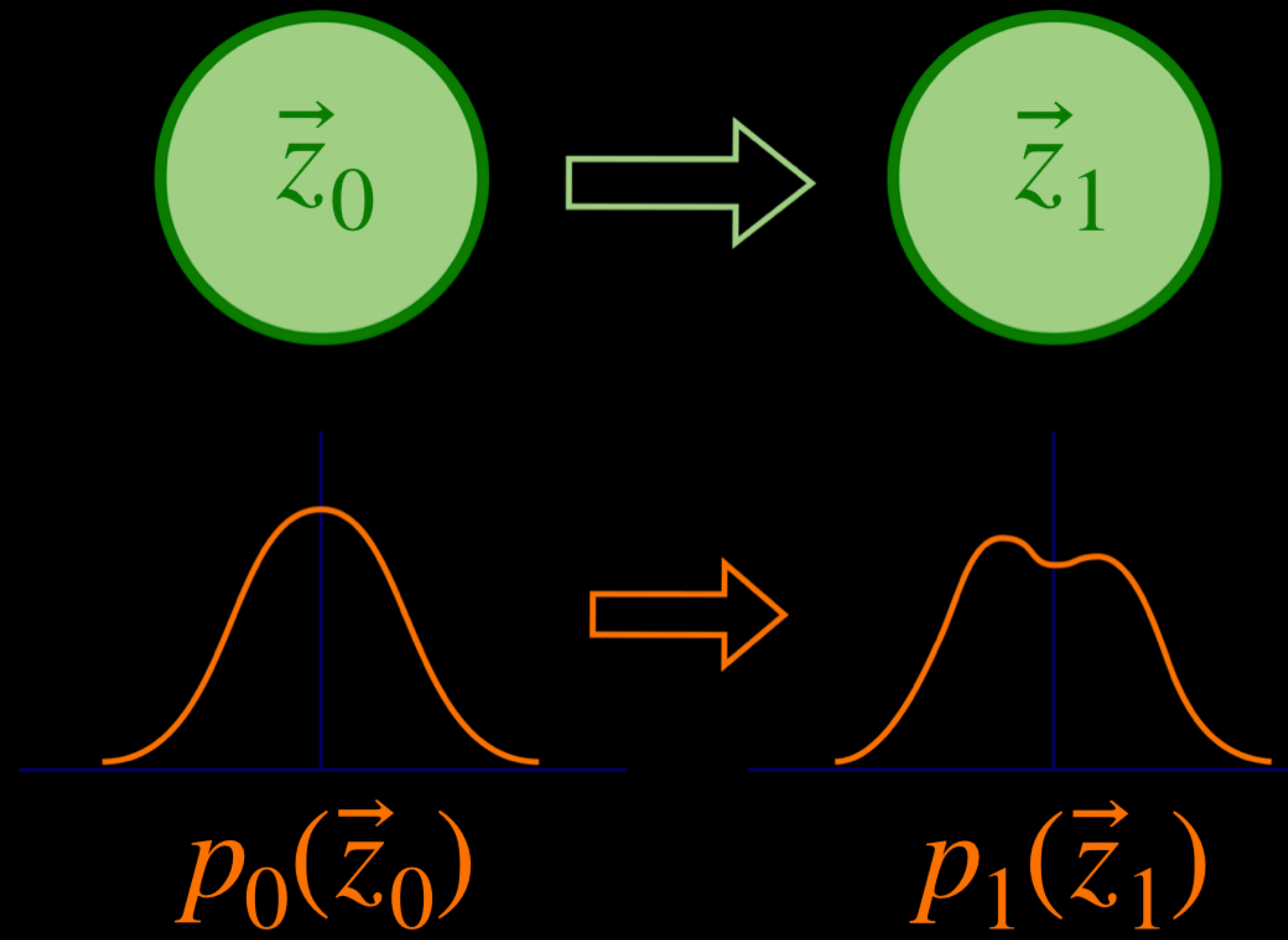
$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log\left|\det\frac{\partial \vec{f}}{\partial \vec{z}_0}\right|$$

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$



$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

Rezende, D.J. and Mohamed, S., 2015. Variational inference with normalizing flows

Normalizing flow

$$\vec{z}_K = f_K \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$$

Normalizing flow

$$\vec{z}_K = f_K \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$$

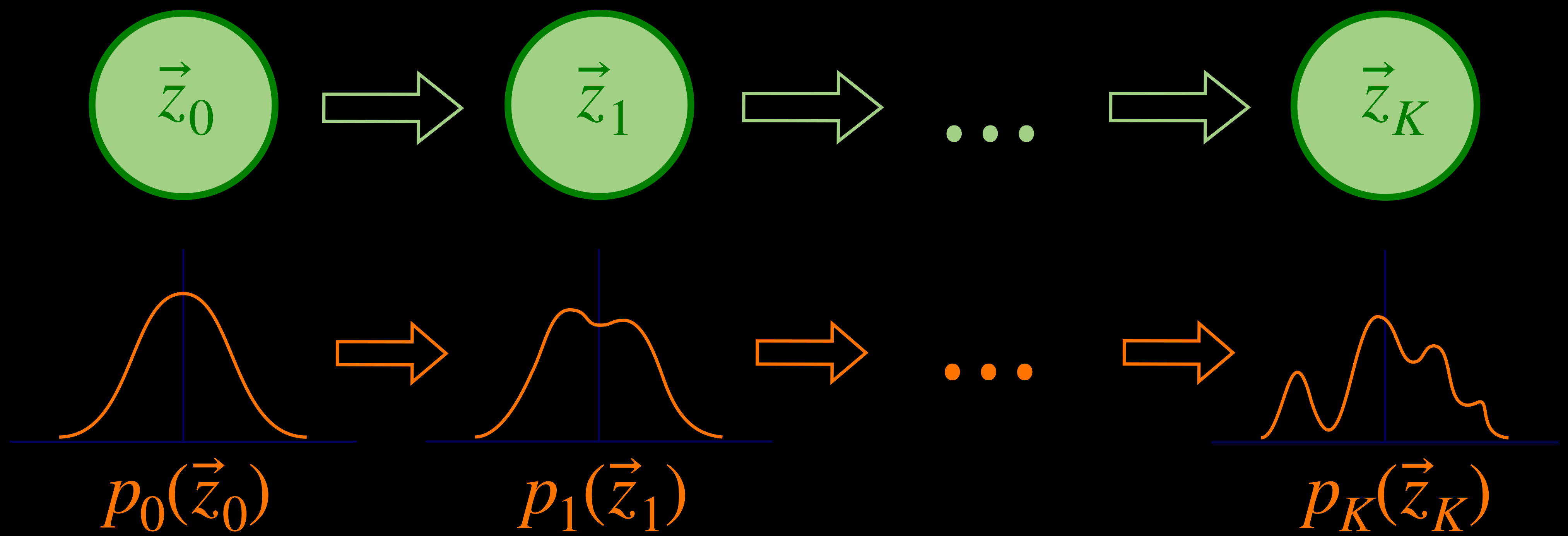
$$\log p_K(\vec{z}_K) = \log p_0(\vec{z}_0) - \sum_{i=1}^{K} \log \left| \det \frac{\partial \overrightarrow{f_i}}{\partial \vec{z}_{i-1}} \right|$$

Normalizing flow

$$\vec{z}_K = f_K \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$$

$$\log p_K(\vec{z}_K) = \log p_0(\vec{z}_0) - \sum_{i=1}^{K} \log \left| \det \frac{\partial \overrightarrow{f_i}}{\partial \vec{z}_{i-1}} \right|$$

Distribution flows through a sequence of invertible transformations

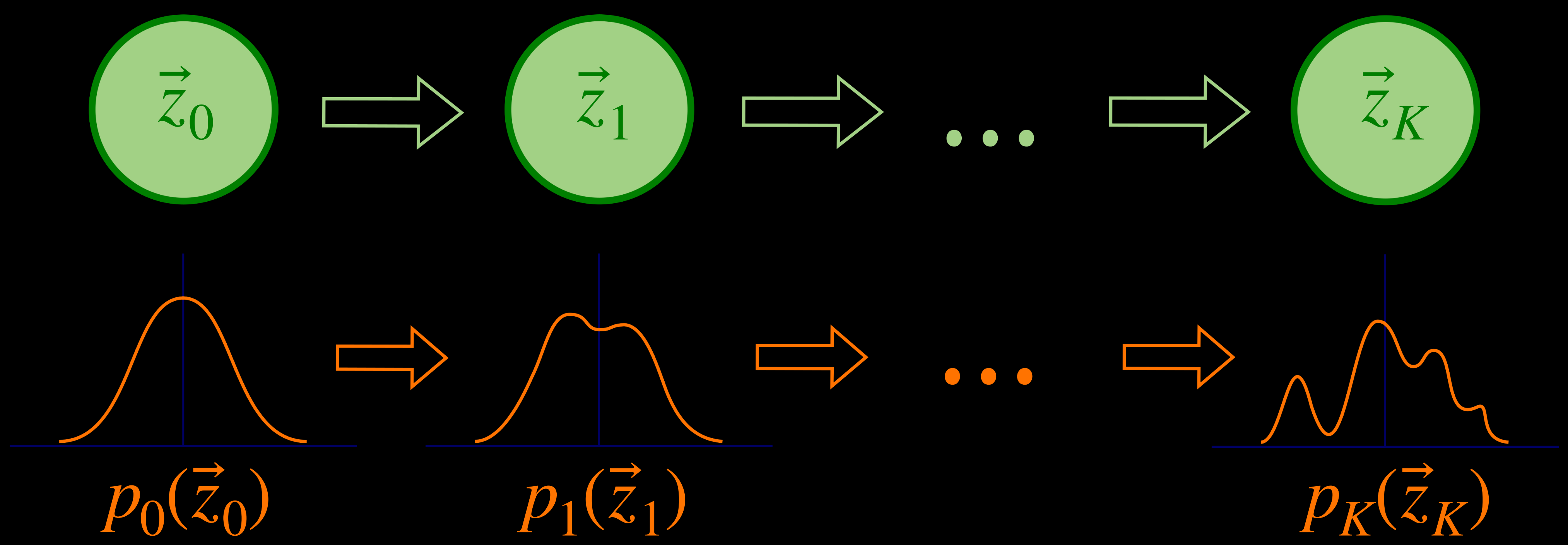


Normalizing flow

$$\vec{z}_K = f_K \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$$

$$\log p_K(\vec{z}_K) = \log p_0(\vec{z}_0) - \boxed{\sum_{i=1}^{K} \log \left| \det \frac{\partial \overrightarrow{f_i}}{\partial \vec{z}_{i-1}} \right|} \qquad O(D^3)$$

Distribution flows through a sequence of invertible transformations



Continuous normalizing flow

Continuous normalizing flow

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

Chen, Tian Qi, et al. "Neural ordinary differential equations."

Continuous normalizing flow

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$\vec{z}(0) \sim p_0(\vec{z}(0))$$

Chen, Tian Qi, et al. "Neural ordinary differential equations."

Continuous normalizing flow

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$\vec{z}(0) \sim p_0(\vec{z}(0))$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), t)dt$$

Chen, Tian Qi, et al. "Neural ordinary differential equations."

Continuous normalizing flow

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$\vec{z}(0) \sim \underline{p_0(\vec{z}(0))}$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), t)dt$$

Chen, Tian Qi, et al. "Neural ordinary differential equations."

Continuous normalizing flow

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$\vec{z}(0) \sim \underline{p_0(\vec{z}(0))}$$

$$\vec{z}(T) = \vec{z}(0) + \int_0^T \vec{f}(\vec{z}(t), t)dt$$

$$\frac{d \log p(\vec{z}(t))}{dt} = ?$$

Chen, Tian Qi, et al. "Neural ordinary differential equations."

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau) d\tau$$

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau)\, d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t))$$

transformation of z

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau) d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t))$$

invertible transformation of z

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t))$$

invertible
transformation of z

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|$$

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t))$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|}{\epsilon}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau) d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t), t) \cdot \epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|$$

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = - \lim_{\epsilon \to 0} \frac{\log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|}{\epsilon}$$

$$\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = I + \frac{\partial \vec{f}(\vec{z}(t), t)}{\partial \vec{z}(t)} \epsilon + o(\epsilon)$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$
$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right|}{\epsilon}$$

$$\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = I + \frac{\partial \vec{f}(\vec{z}(t),t)}{\partial \vec{z}(t)}\epsilon + o(\epsilon) \to I$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau) d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t), t) \cdot \epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|$$

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon \to 0} \frac{\log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|}{\epsilon}$$

$$= -\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} \log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|$$

$$\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = I + \frac{\partial \vec{f}(\vec{z}(t), t)}{\partial \vec{z}(t)} \epsilon + o(\epsilon) \to I$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|}{\epsilon}$$

$$= -\lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)} = I + \frac{\partial\vec{f}(\vec{z}(t),t)}{\partial\vec{z}(t)}\epsilon + o(\epsilon)$$

$$\frac{\partial}{\partial\epsilon}\log\left|\det A(\epsilon)\right| = tr\left[A^{-1}\frac{\partial}{\partial\epsilon}A\right]$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$
$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|}{\epsilon}$$

$$= -\lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right| = -\lim_{\epsilon\to 0}\ tr\Big[\Big(\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\Big)^{-1}\frac{\partial}{\partial\epsilon}\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\Big]$$

$$\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)} = I + \frac{\partial\vec{f}(\vec{z}(t),t)}{\partial\vec{z}(t)}\epsilon + o(\epsilon)$$

$$\frac{\partial}{\partial\epsilon}\log\Big|\det A(\epsilon)\Big| = tr\Big[A^{-1}\frac{\partial}{\partial\epsilon}A\Big]$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$
$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|}{\epsilon}$$

$$= -\lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right| = -\lim_{\epsilon\to 0}\ tr\left[\left(\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right)^{-1}\frac{\partial}{\partial\epsilon}\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right]$$

$$\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)} = I + \frac{\partial\vec{f}(\vec{z}(t),t)}{\partial\vec{z}(t)}\epsilon + o(\epsilon) \qquad \frac{\partial}{\partial\epsilon}\log\left|\det A(\epsilon)\right| = tr\left[A^{-1}\frac{\partial}{\partial\epsilon}A\right]$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right|}{\epsilon}$$

$$= -\lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\log\left|\det\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right| = -\lim_{\epsilon\to 0}\ tr\left[\left(\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right)^{-1}\frac{\partial}{\partial\epsilon}\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right]$$

$$\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)} = I + \frac{\partial\vec{f}(\vec{z}(t),t)}{\partial\vec{z}(t)}\epsilon + o(\epsilon)$$

$$\lim_{\epsilon\to 0}\left(\frac{\partial\overrightarrow{T}_\epsilon}{\partial\vec{z}(t)}\right)^{-1} = I$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau), \tau) d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t), t) \cdot \epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|$$

$$\frac{d \log p(\vec{z})}{dt} = \lim_{\epsilon \to 0} \frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon \to 0} \frac{\log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right|}{\epsilon}$$

$$= -\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} \log \left| \det \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right| = -\lim_{\epsilon \to 0} tr\left[\left(\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right)^{-1} \frac{\partial}{\partial \epsilon} \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right]$$

$$\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = I + \frac{\partial \vec{f}(\vec{z}(t), t)}{\partial \vec{z}(t)} \epsilon + o(\epsilon)$$

$$\lim_{\epsilon \to 0} \left(\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} \right)^{-1} = I \qquad \lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} \frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = \frac{\partial \vec{f}(\vec{z}(t), t)}{\partial \vec{z}(t)}$$

$$\vec{z}(t+\epsilon) = \vec{z}(t) + \int_t^{t+\epsilon} \vec{f}(\vec{z}(\tau),\tau)d\tau = \overrightarrow{T}_\epsilon(\vec{z}(t)) =$$

$$= \vec{z}(t) + \vec{f}(\vec{z}(t),t)\cdot\epsilon + o(\epsilon)$$

$$\log p(\vec{z}(t+\epsilon)) = \log p(\vec{z}(t)) - \log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right|$$

$$\frac{d\log p(\vec{z})}{dt} = \lim_{\epsilon\to 0}\frac{\log p(\vec{z}(t+\epsilon)) - \log p(\vec{z}(t))}{\epsilon} = -\lim_{\epsilon\to 0}\frac{\log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right|}{\epsilon}$$

$$= -\lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\log\left|\det\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right| = -\lim_{\epsilon\to 0}\ tr\Big[\left(\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right)^{-1}\frac{\partial}{\partial\epsilon}\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\Big] =$$

$$\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = I + \frac{\partial \vec{f}(\vec{z}(t),t)}{\partial \vec{z}(t)}\epsilon + o(\epsilon)$$

$$= -\,tr\Big[\frac{\partial \vec{f}(\vec{z}(t),t)}{\partial \vec{z}(t)}\Big]$$

$$\lim_{\epsilon\to 0}\left(\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)}\right)^{-1} = I \qquad\qquad \lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\frac{\partial \overrightarrow{T}_\epsilon}{\partial \vec{z}(t)} = \frac{\partial \vec{f}(\vec{z}(t),t)}{\partial \vec{z}(t)}$$

Continuous normalizing flow

$$\vec{z}(0) \sim p(\vec{z}(0))$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$\boxed{\frac{d \log p(\vec{z}(t))}{dt} = -\, tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)}$$

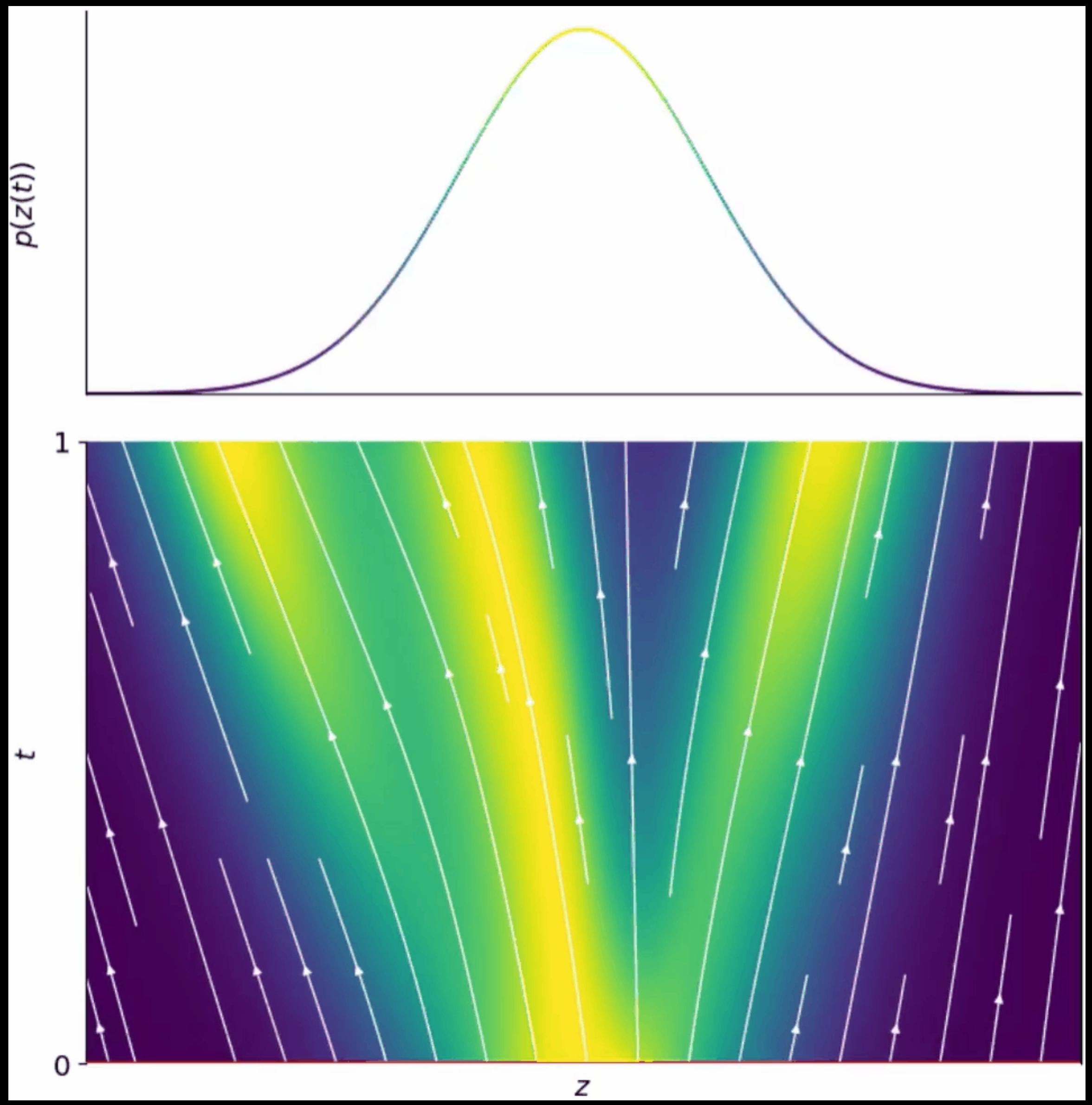
Chen, Tian Qi, et al. "Neural ordinary differential equations."

Continuous normalizing flow

$$\vec{z}(0) \sim p(\vec{z}(0))$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$\boxed{\frac{d \log p(\vec{z}(t))}{dt} = -\, tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)}$$



Chen, Tian Qi, et al. "Neural ordinary differential equations."

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ invertible and smooth

$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

Continuous normalizing flow

$$\vec{z}(t_0) = \vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$\vec{f}$ uniformly Lipschitz continuous in z and continuous in t

$$\frac{d \log p(\vec{z}(t))}{dt} = -tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)$$

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ invertible and smooth

$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

$O(D^3)$ for det computation

Continuous normalizing flow

$$\vec{z}(t_0) = \vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$\vec{f}$ uniformly Lipschitz continuous in z and continuous in t

$$\frac{d \log p(\vec{z}(t))}{dt} = -\,tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)$$

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ invertible and smooth

$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

$O(D^3)$ for det computation

Continuous normalizing flow

$$\vec{z}(t_0) = \vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$\vec{f}$ uniformly Lipschitz continuous in z and continuous in t

$$\frac{d \log p(\vec{z}(t))}{dt} = -\, tr\left(\frac{\partial \vec{f}}{\partial \vec{z}} \right)$$

$O(D)$ for trace computation?

Normalizing flow

Continuous normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ invertible and smooth

$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

$O(D^3)$ for det computation

$$\vec{z}(t_0) = \vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$\vec{f}$ uniformly Lipschitz continuous in z and continuous in t

$$\frac{d \log p(\vec{z}(t))}{dt} = -\, tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)$$

$O(D)$ for trace computation?

$$tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right) = \frac{\partial f_1}{\partial z_1} + \frac{\partial f_2}{\partial z_2} + \ldots + \frac{\partial f_D}{\partial z_D}$$

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ invertible and smooth

$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

$O(D^3)$ for det computation

Continuous normalizing flow

$$\vec{z}(t_0) = \vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$\vec{f}$ uniformly Lipschitz continuous in z and continuous in t

$$\frac{d \log p(\vec{z}(t))}{dt} = -\,tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)$$

$O(D)$ for trace computation?

$$tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right) = \underbrace{\frac{\partial f_1}{\partial z_1} + \frac{\partial f_2}{\partial z_2} + \ldots + \frac{\partial f_D}{\partial z_D}}_{\text{D terms}}$$

Normalizing flow

$$\vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\vec{z}_1 = \vec{f}(\vec{z}_0)$$

$\vec{f}$ invertible and smooth

$$\log p_1(\vec{z}_1) = \log p_0(\vec{z}_0) - \log \left| \det \frac{\partial \vec{f}}{\partial \vec{z}_0} \right|$$

$O(D^3)$ for det computation

Continuous normalizing flow

$$\vec{z}(t_0) = \vec{z}_0 \sim p_0(\vec{z}_0)$$

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$\vec{f}$ uniformly Lipschitz continuous in z and continuous in t

$$\frac{d \log p(\vec{z}(t))}{dt} = -\,tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right)$$

$O(D)$ for trace computation?

$$tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right) = \underbrace{\frac{\partial f_1}{\partial z_1} + \frac{\partial f_2}{\partial z_2} + \ldots + \frac{\partial f_D}{\partial z_D}}_{\text{D terms}}$$

$$costs\left(\frac{\partial f_i}{\partial z_i}\right) \simeq costs(\vec{f})$$

Automatic differentiation

$$f_1 = z_1 z_2 + \sin z_1 \qquad\qquad f_2 = z_1 z_2 \sin z_1$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad\qquad f_2 = z_1 z_2 \sin z_1$$

Automatic differentiation

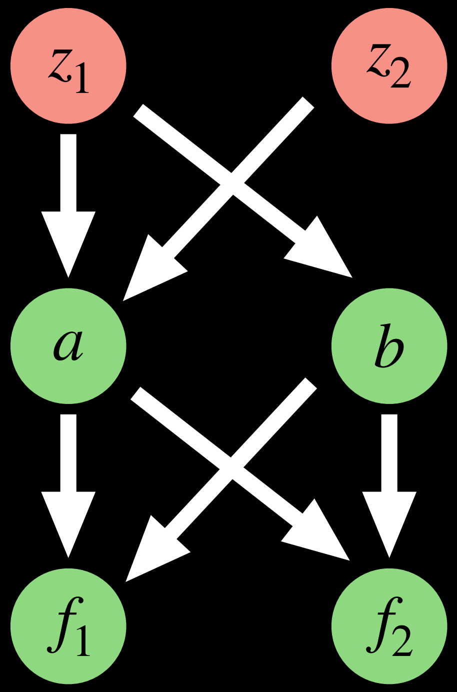
$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

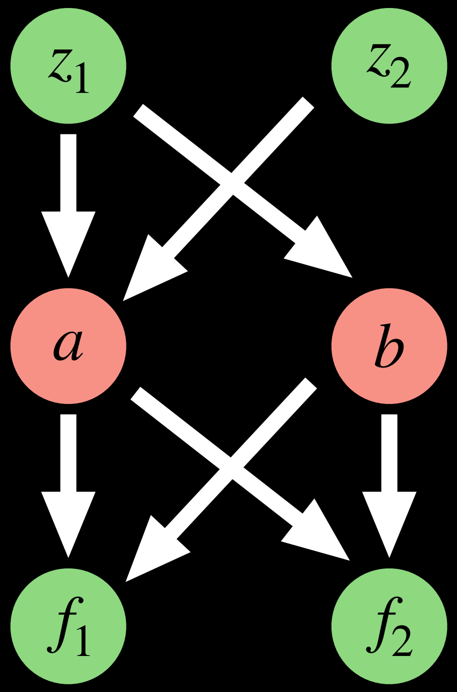


z_1 z_2

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



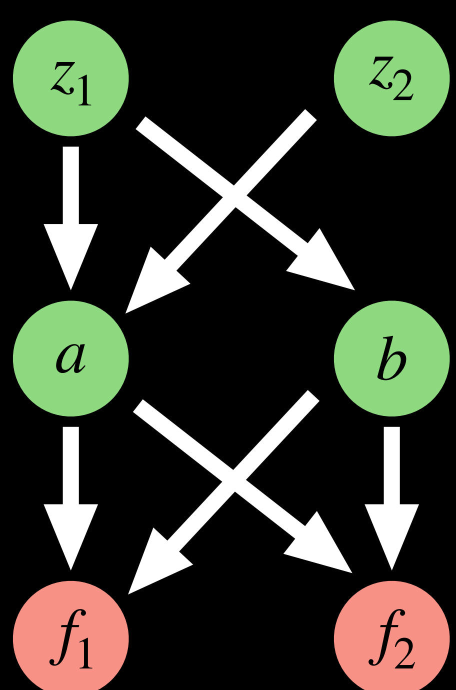
$z_1 \qquad z_2$

$a = z_1 z_2 \qquad b = \sin z_1$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



z_1 $\quad$ z_2

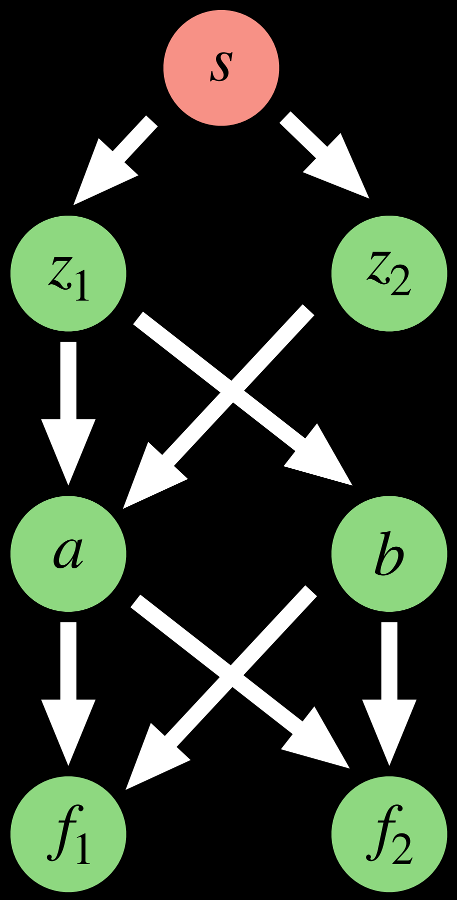
$a = z_1 z_2 \quad b = \sin z_1$

$f_1 = a + b \quad f_2 = a \cdot b$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

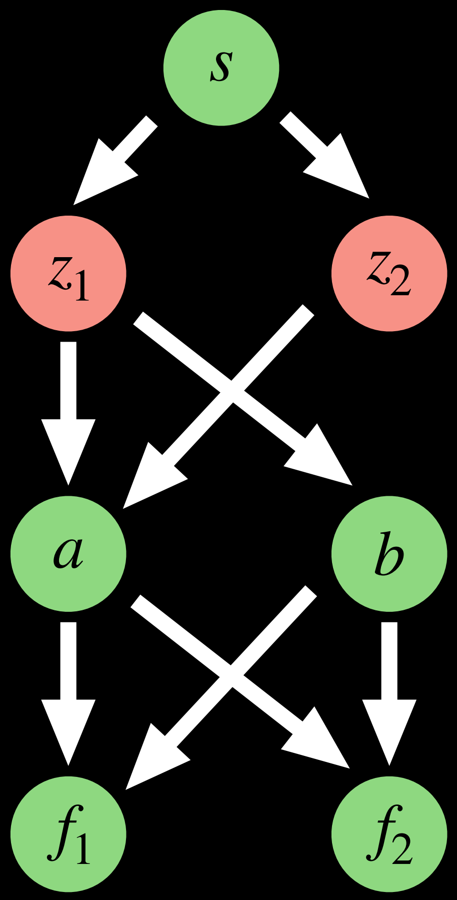


Forward-mode automatic differentiation

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



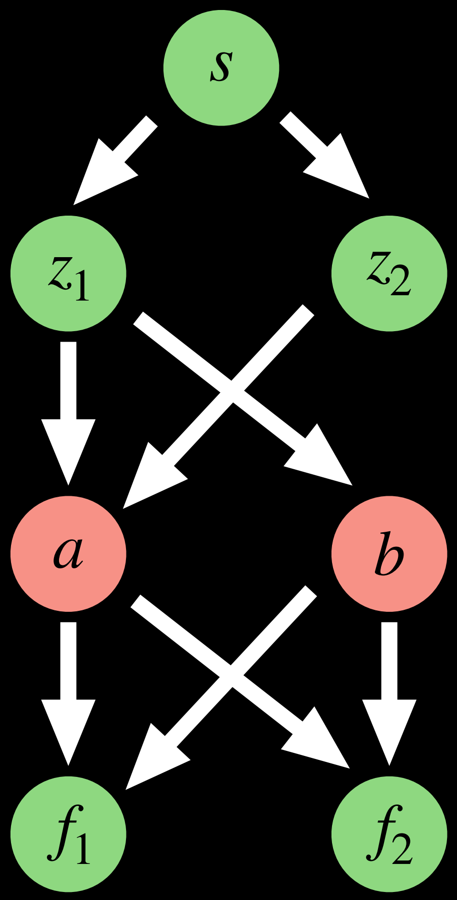
Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} \qquad \frac{\partial z_2}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

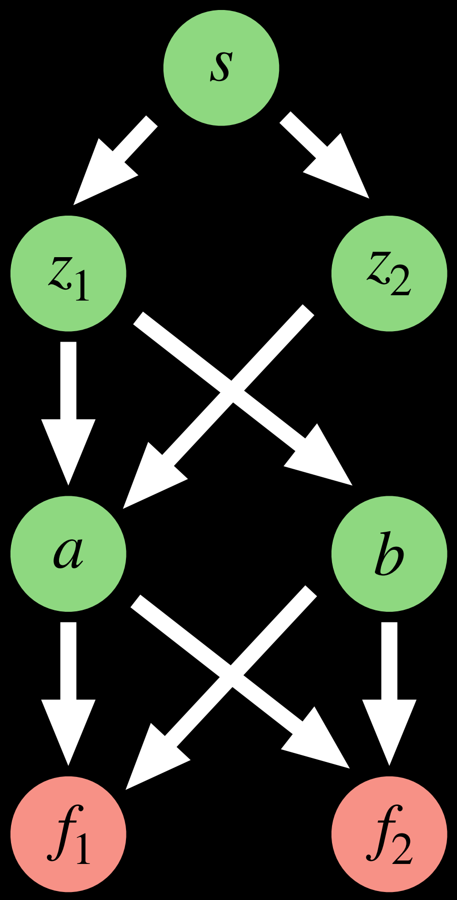
$$\frac{\partial z_1}{\partial s} \qquad \frac{\partial z_2}{\partial s}$$

$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} \qquad \frac{\partial z_2}{\partial s}$$

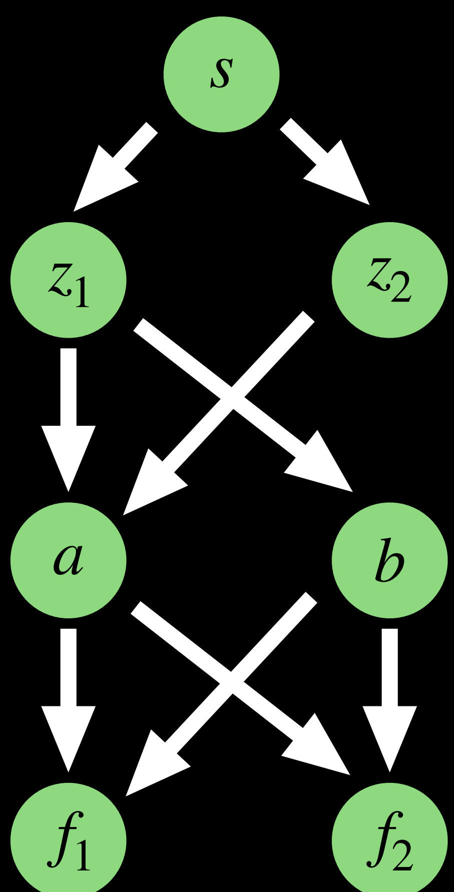
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial s} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial s} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} \qquad \frac{\partial z_2}{\partial s} \qquad s = z_1$$

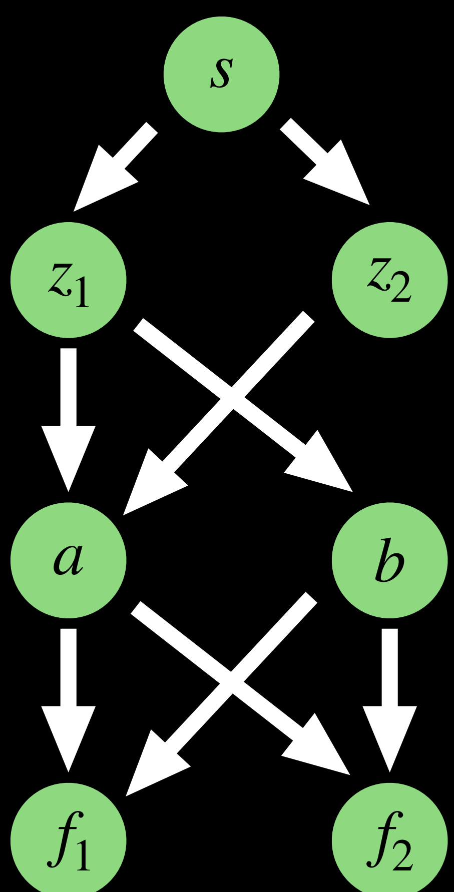
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial s} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial s} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} = 1 \qquad \frac{\partial z_2}{\partial s} = 0 \qquad s = z_1$$

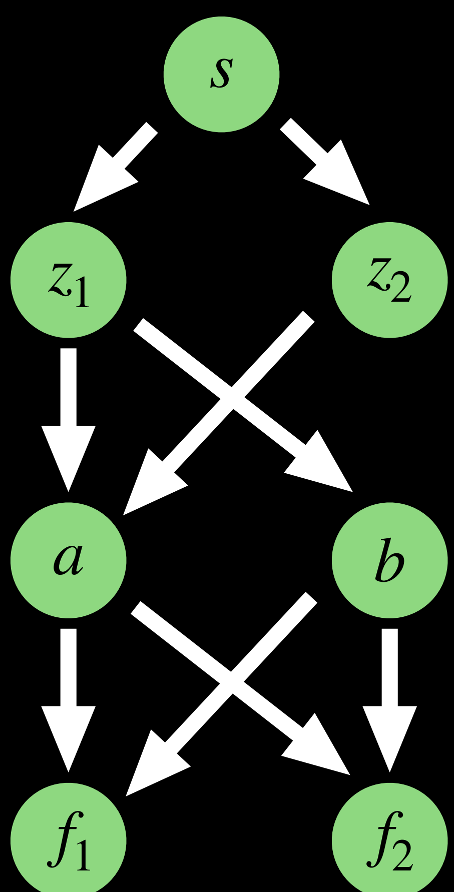
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial s} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial s} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} = 1 \qquad \frac{\partial z_2}{\partial s} = 0 \qquad s = z_1$$

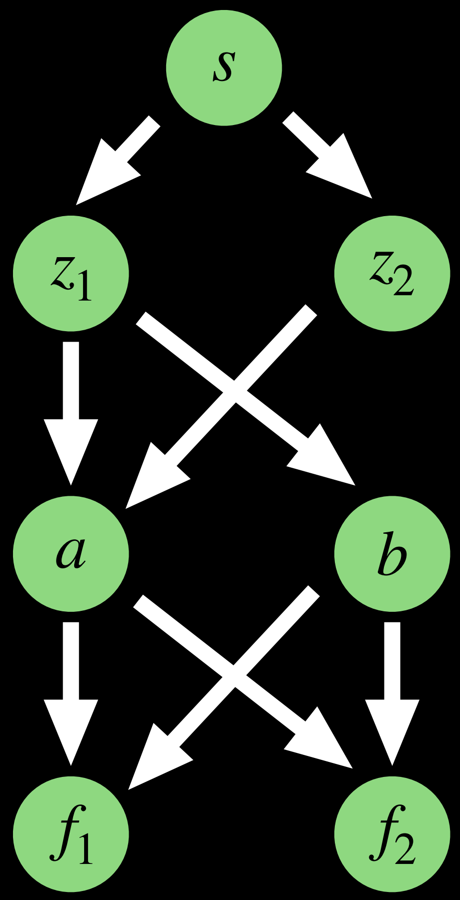
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial z_1} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial z_1} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} = 0 \qquad \frac{\partial z_2}{\partial s} = 1 \qquad s = z_2$$

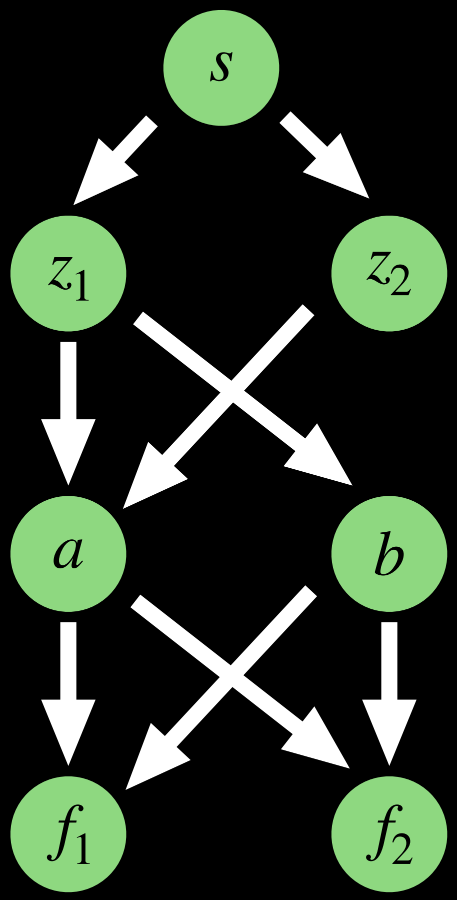
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial s} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial s} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} = 0 \qquad \frac{\partial z_2}{\partial s} = 1 \qquad s = z_2$$

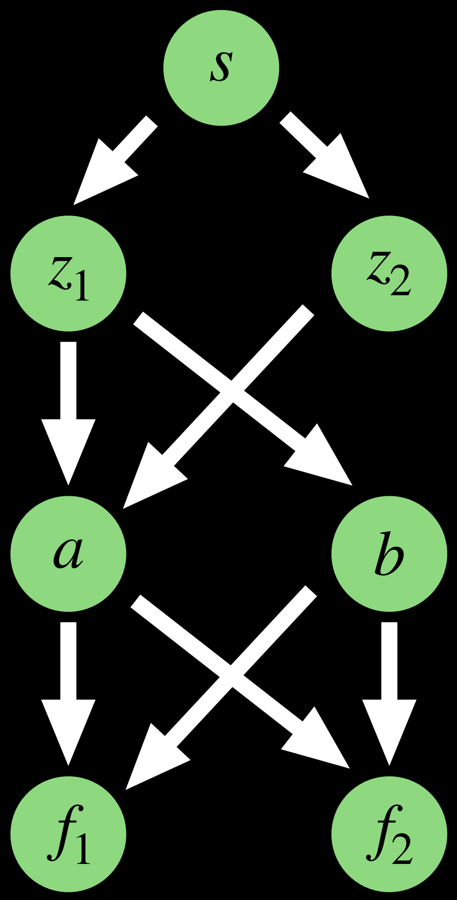
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial z_2} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial z_2} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} \qquad \frac{\partial z_2}{\partial s} \qquad s = z_1 \;\; or \;\; s = z_2$$

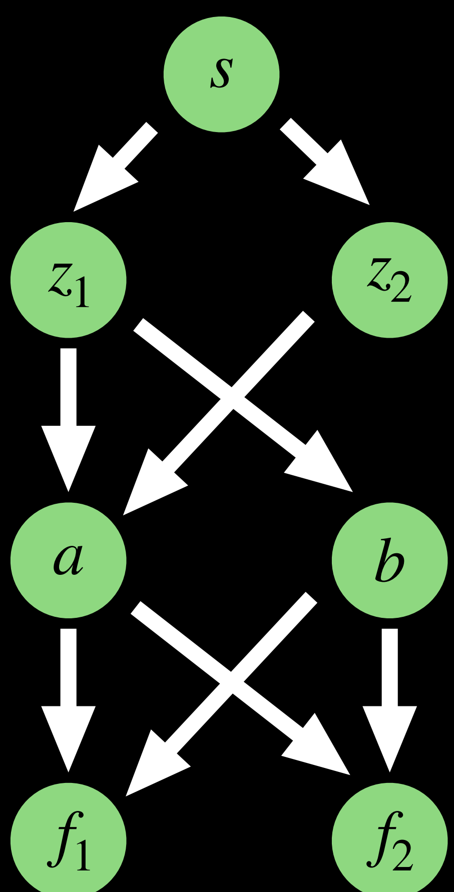
$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial s} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial s} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Forward-mode automatic differentiation

$$\frac{\partial z_1}{\partial s} \qquad \frac{\partial z_2}{\partial s} \qquad s = z_1 \;\; or \;\; s = z_2$$

$$\frac{\partial a}{\partial s} = z_1 \frac{\partial z_2}{\partial s} + z_2 \frac{\partial z_1}{\partial s} \qquad \frac{\partial b}{\partial s} = \cos z_1 \frac{\partial z_1}{\partial s}$$

$$\frac{\partial f_1}{\partial s} = \frac{\partial a}{\partial s} + \frac{\partial b}{\partial s} \qquad \frac{\partial f_2}{\partial s} = a \frac{\partial b}{\partial s} + b \frac{\partial a}{\partial s}$$

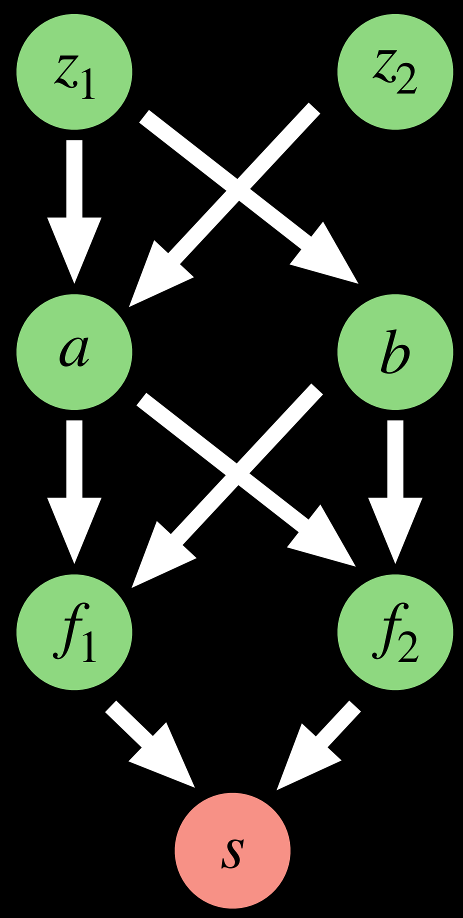
$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs(\vec{f})$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation

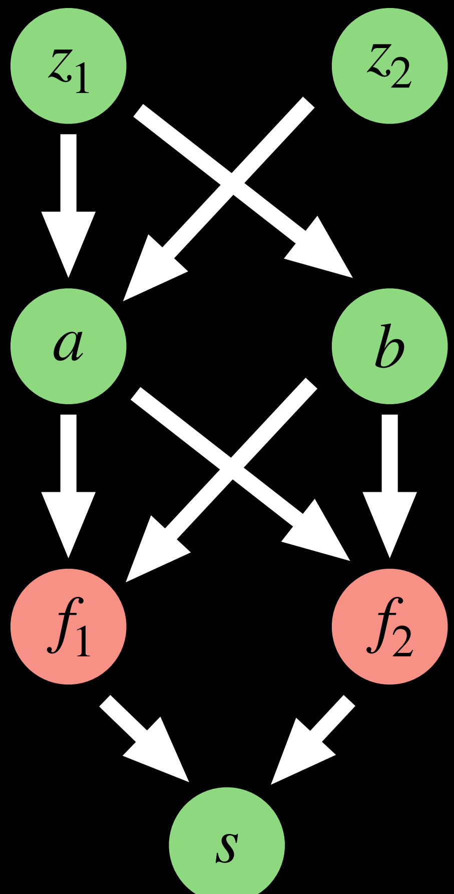


Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



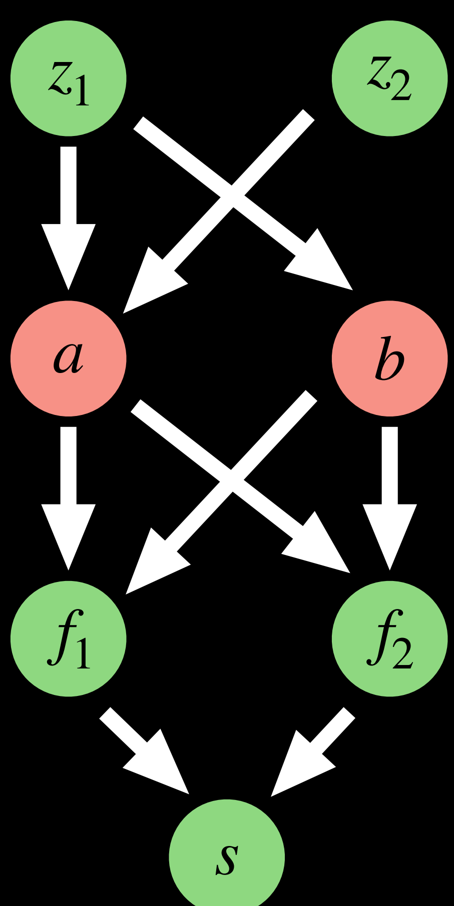
$$\frac{\partial s}{\partial f_1} \qquad \frac{\partial s}{\partial f_2}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

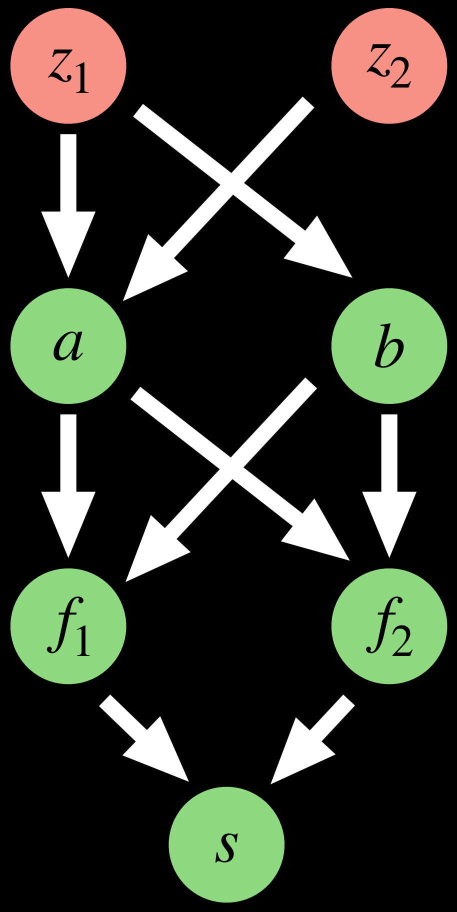
$$\frac{\partial s}{\partial f_1} \qquad \frac{\partial s}{\partial f_2}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



$$\frac{\partial s}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad \frac{\partial s}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

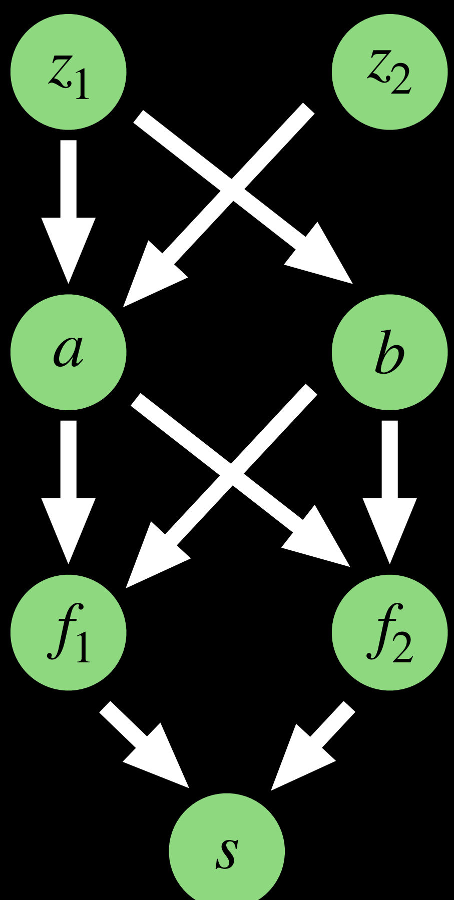
$$\frac{\partial s}{\partial f_1} \qquad \frac{\partial s}{\partial f_2}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



$$\frac{\partial s}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad \frac{\partial s}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

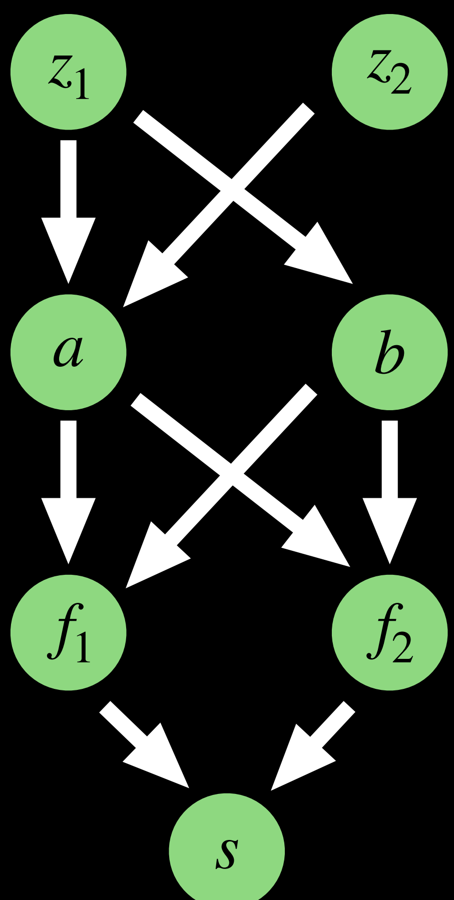
$$\frac{\partial s}{\partial f_1} = 1 \qquad \frac{\partial s}{\partial f_2} = 0 \qquad s = f_1$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



$$\frac{\partial f_1}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad \frac{\partial f_1}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

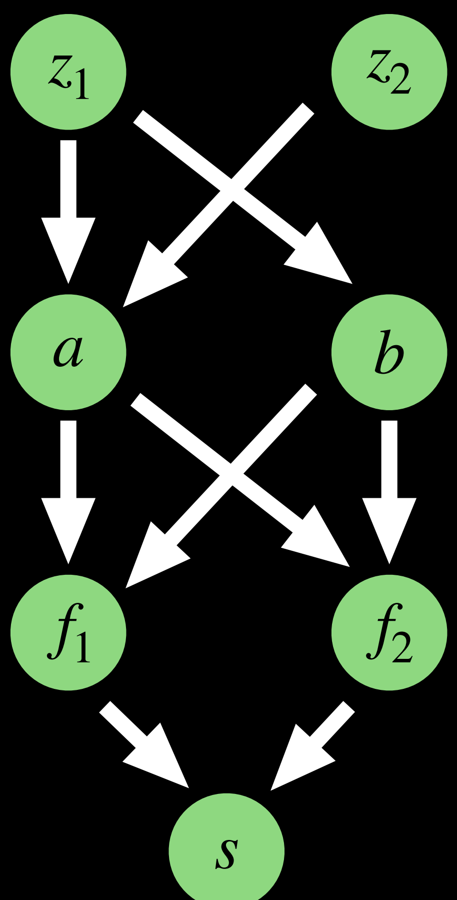
$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

$$\frac{\partial s}{\partial f_1} = 1 \qquad \frac{\partial s}{\partial f_2} = 0 \qquad s = f_1$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Reverse-mode automatic differentiation

$$\frac{\partial s}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad \frac{\partial s}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

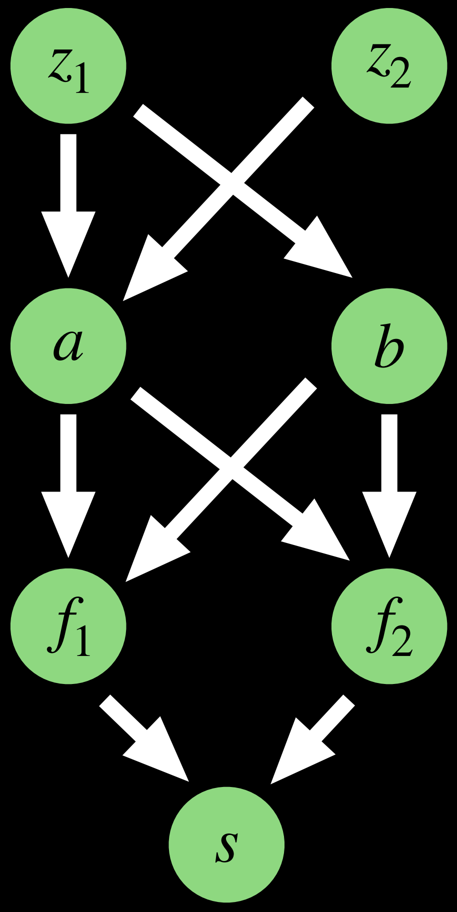
$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

$$\frac{\partial s}{\partial f_1} = 0 \qquad \frac{\partial s}{\partial f_2} = 1 \qquad s = f_2$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



Reverse-mode automatic differentiation

$$\frac{\partial f_2}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad \frac{\partial f_2}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

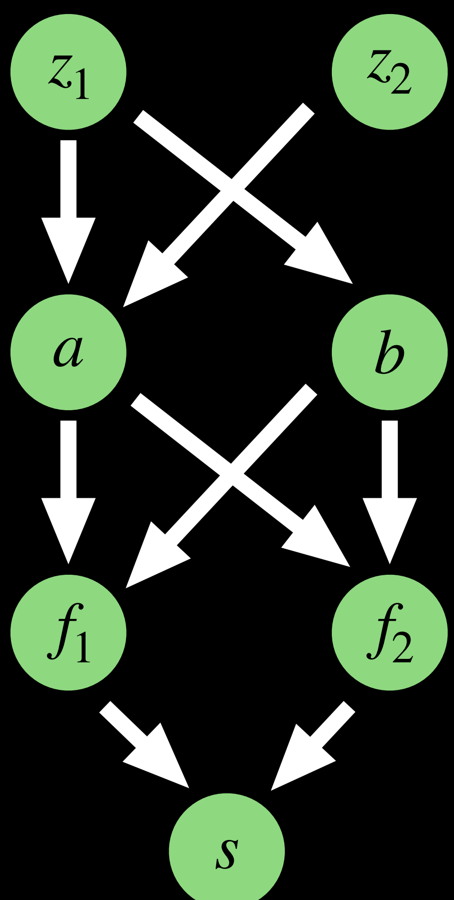
$$\frac{\partial s}{\partial f_1} = 0 \qquad \frac{\partial s}{\partial f_2} = 1 \qquad s = f_2$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad\qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad\qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



$$\frac{\partial s}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad\qquad \frac{\partial s}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad\qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

$$\frac{\partial s}{\partial f_1} \qquad \frac{\partial s}{\partial f_2}$$

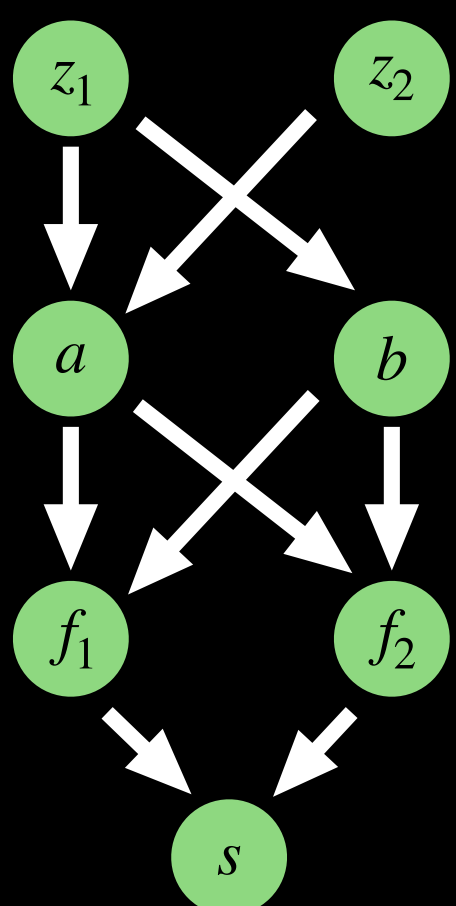
$$s = f_1 \;\; or \;\; s = f_2$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

Reverse-mode automatic differentiation



$$\frac{\partial s}{\partial z_1} = \frac{\partial s}{\partial a} z_2 + \frac{\partial s}{\partial b} cos(z_1) \qquad \frac{\partial s}{\partial z_2} = \frac{\partial s}{\partial a} z_1$$

$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} b \qquad \frac{\partial s}{\partial b} = \frac{\partial s}{\partial f_1} + \frac{\partial s}{\partial f_2} a$$

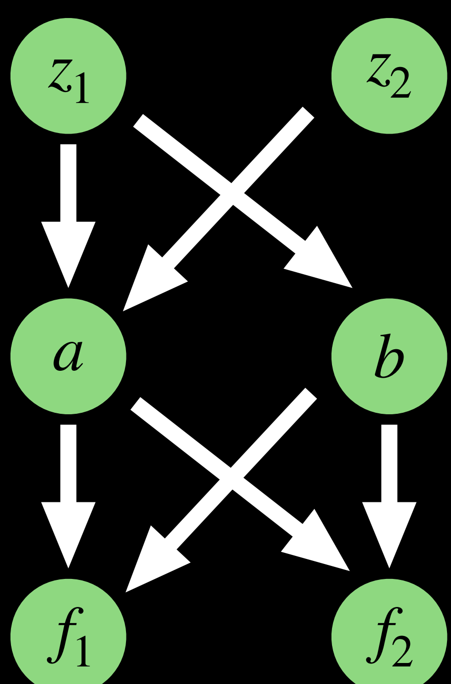
$$\frac{\partial s}{\partial f_1} \qquad \frac{\partial s}{\partial f_2} \qquad s = f_1 \;\; or \;\; s = f_2$$

$$costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$

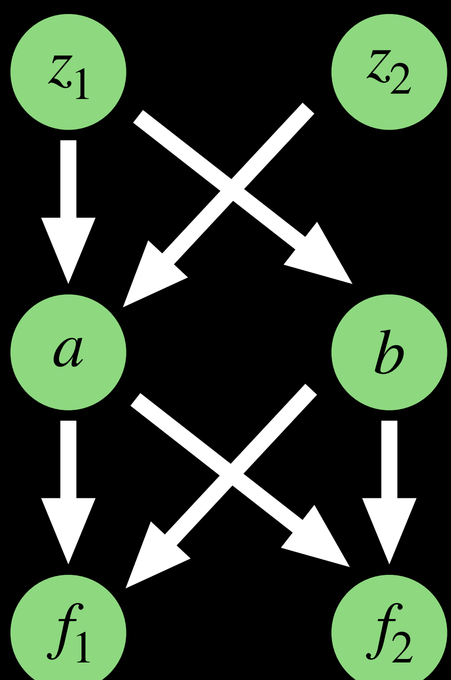


$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



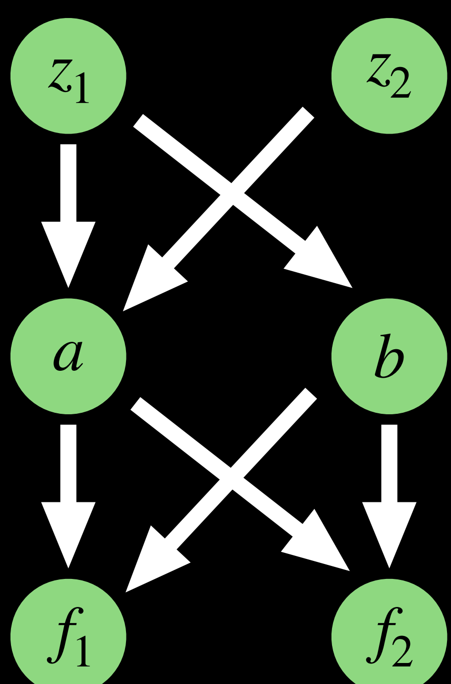
$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

$$tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right) = \frac{\partial f_1}{\partial z_1} + \frac{\partial f_2}{\partial z_2} + \ldots + \frac{\partial f_D}{\partial z_D}$$

Automatic differentiation

$$f_1 = \underbrace{z_1 z_2}_{a} + \underbrace{\sin z_1}_{b} \qquad f_2 = z_1 z_2 \sin z_1$$

$$f_1 = a + b \qquad f_2 = a \cdot b$$



$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

$$tr\left(\frac{\partial \vec{f}}{\partial \vec{z}}\right) = \frac{\partial f_1}{\partial z_1} + \frac{\partial f_2}{\partial z_2} + \ldots + \frac{\partial f_D}{\partial z_D}$$

$$O(D \cdot costs(\vec{f}))$$

Free-form Jacobian of Reversible Dynamics

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

1. $\overrightarrow{v}^T \dfrac{\partial \vec{f}}{\partial \vec{z}}$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

1. $$\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}} = \frac{\partial}{\partial \vec{z}}(\overrightarrow{v}^T \vec{f})$$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

1. $\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}} = \frac{\partial}{\partial \vec{z}}(\overrightarrow{v}^T \vec{f})$ $\quad costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

1. $\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}} = \frac{\partial}{\partial \vec{z}}(\overrightarrow{v}^T \vec{f}) \quad costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $\mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}]$

$\mathbb{E}(\overrightarrow{\epsilon}) = 0$ and $Cov(\overrightarrow{\epsilon}) = I$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

1. $\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}} = \frac{\partial}{\partial \vec{z}}(\overrightarrow{v}^T \vec{f})$ $\quad costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $\mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}] = \mathbb{E}_{p(\overrightarrow{\epsilon})}\left[\sum_{i,j} \epsilon_i \cdot a_{ij} \cdot \epsilon_j\right]$

$\mathbb{E}(\overrightarrow{\epsilon}) = 0$ and $Cov(\overrightarrow{\epsilon}) = I$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

$$costs\left(\frac{\partial \vec{f}}{\partial z_i}\right) \simeq costs\left(\frac{\partial f_i}{\partial \vec{z}}\right) \simeq costs(\vec{f})$$

1. $\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}} = \frac{\partial}{\partial \vec{z}}(\overrightarrow{v}^T \vec{f}) \quad costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $\mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}] = \mathbb{E}_{p(\overrightarrow{\epsilon})}\left[\sum_{i,j} \epsilon_i \cdot a_{ij} \cdot \epsilon_j\right] = \sum_{i,j} \delta_{ij} \cdot a_{ij} = tr(A)$

$\mathbb{E}(\overrightarrow{\epsilon}) = 0$ and $Cov(\overrightarrow{\epsilon}) = I$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

1. $costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $tr(A) = \mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}]$ $\mathbb{E}(\overrightarrow{\epsilon}) = 0$ and $Cov(\overrightarrow{\epsilon}) = I$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

1. $costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $tr(A) = \mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}]$ $\mathbb{E}(\overrightarrow{\epsilon}) = 0$ **and** $Cov(\overrightarrow{\epsilon}) = I$

$$\log p(\vec{z}(T)) = \log p(\vec{z}(0)) - \int_0^T tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt$$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

1. $costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $tr(A) = \mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}]$ $\qquad$ $\mathbb{E}(\overrightarrow{\epsilon}) = 0$ **and** $Cov(\overrightarrow{\epsilon}) = I$

$$\log p(\vec{z}(T)) = \log p(\vec{z}(0)) - \int_0^T tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt$$

$$= \log p(\vec{z}(0)) - \int_0^T \mathbb{E}_{p(\overrightarrow{\epsilon})}\left[\overrightarrow{\epsilon}^T \frac{\partial f}{\partial \vec{z}} \overrightarrow{\epsilon}\right] dt$$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Free-form Jacobian of Reversible Dynamics (FFJORD)

1. $costs\left(\overrightarrow{v}^T \frac{\partial \vec{f}}{\partial \vec{z}}\right) \simeq costs(\vec{f})$

2. $tr(A) = \mathbb{E}_{p(\overrightarrow{\epsilon})}[\overrightarrow{\epsilon}^T A \overrightarrow{\epsilon}]$ $\mathbb{E}(\overrightarrow{\epsilon}) = 0$ and $Cov(\overrightarrow{\epsilon}) = I$

$$\log p(\vec{z}(T)) = \log p(\vec{z}(0)) - \int_0^T tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt$$

$$= \log p(\vec{z}(0)) - \int_0^T \mathbb{E}_{p(\overrightarrow{\epsilon})}\left[\overrightarrow{\epsilon}^T \frac{\partial f}{\partial \vec{z}} \overrightarrow{\epsilon}\right] dt$$

$$= \log p(\vec{z}(0)) - \mathbb{E}_{p(\overrightarrow{\epsilon})}\left[\int_0^T \overrightarrow{\epsilon}^T \frac{\partial f}{\partial \vec{z}} \overrightarrow{\epsilon}\, dt\right]$$

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Likelihood computational cost

NF: $\vec{z}_L = f_L \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$

$$\log p_L(\vec{z}_L) = \log p_0(\vec{z}_0) - \sum_{i=1}^{L} \log \left| \det \frac{\partial \vec{f_i}}{\partial \vec{z}_{i-1}} \right|$$

Likelihood computational cost

NF: $\vec{z}_L = f_L \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$ $O(D^3 \cdot L)$

$$\log p_L(\vec{z}_L) = \log p_0(\vec{z}_0) - \underline{\sum_{i=1}^{L} \log \left| \det \frac{\partial \vec{f_i}}{\partial \vec{z}_{i-1}} \right|}$$

Likelihood computational cost

NF: $\vec{z}_L = f_L \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$ $\quad O(D^3 \cdot L)$

$$\log p_L(\vec{z}_L) = \log p_0(\vec{z}_0) - \underline{\sum_{i=1}^{L} \log \left| \det \frac{\partial \vec{f_i}}{\partial \vec{z}_{i-1}} \right|}$$

CNF: $\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$

$$\log p(\vec{z}(t_1)) = \log p(\vec{z}(t_0)) - \int_{t_0}^{t_1} tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt$$

Likelihood computational cost

NF:

$$\vec{z}_L = f_L \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$$

$$O(D^3 \cdot L)$$

$$\log p_L(\vec{z}_L) = \log p_0(\vec{z}_0) - \underline{\sum_{i=1}^{L} \log \left| \det \frac{\partial \overrightarrow{f_i}}{\partial \vec{z}_{i-1}} \right|}$$

CNF:

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$$O(D \cdot cost(\vec{f}) \cdot \hat{L})$$

$$\log p(\vec{z}(t_1)) = \log p(\vec{z}(t_0)) - \underline{\int_{t_0}^{t_1} tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt}$$

Likelihood computational cost

NF:

$$\vec{z}_L = f_L \circ \ldots \circ f_2 \circ f_1(\vec{z}_0)$$

$O(D^3 \cdot L)$

$$\log p_L(\vec{z}_L) = \log p_0(\vec{z}_0) - \underline{\sum_{i=1}^{L} \log \left| \det \frac{\partial \vec{f_i}}{\partial \vec{z}_{i-1}} \right|}$$

CNF:

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t)$$

$O(D \cdot cost(\vec{f}) \cdot \hat{L})$

$$\log p(\vec{z}(t_1)) = \log p(\vec{z}(t_0)) - \underline{\int_{t_0}^{t_1} tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt}$$

FFJORD:

$$\log p(\vec{z}(t_1)) = \log p(\vec{z}(t_0)) - \mathbb{E}_{p(\vec{\epsilon})}\left[\int_{t_0}^{t_1} \vec{\epsilon}^{\,T} \frac{\partial f}{\partial \vec{z}} \vec{\epsilon}\, dt\right]$$

Likelihood computational cost

NF:

$$\vec{z}_L = f_L \circ \ldots \circ f_2 \circ f_1(\vec{z}_0) \qquad O(D^3 \cdot L)$$

$$\log p_L(\vec{z}_L) = \log p_0(\vec{z}_0) - \underline{\sum_{i=1}^{L} \log \left| \det \frac{\partial \vec{f_i}}{\partial \vec{z}_{i-1}} \right|}$$

CNF:

$$\frac{d\vec{z}}{dt} = \vec{f}(\vec{z}(t), t) \qquad O(D \cdot cost(\vec{f}) \cdot \hat{L})$$

$$\log p(\vec{z}(t_1)) = \log p(\vec{z}(t_0)) - \underline{\int_{t_0}^{t_1} tr\left(\frac{\partial f}{\partial \vec{z}}\right) dt}$$

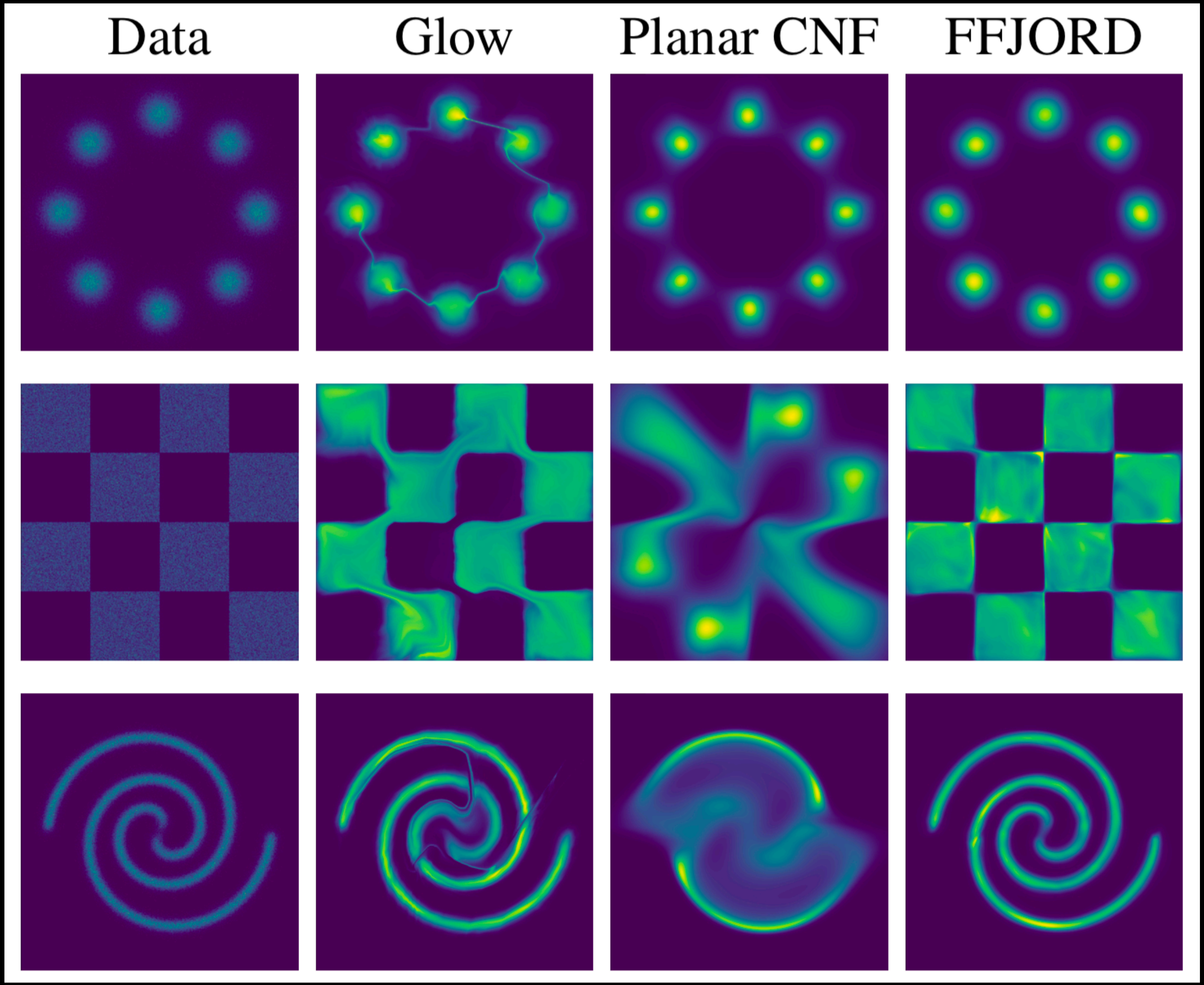
FFJORD:

$$O\Big((cost(\vec{f}) + D) \cdot \hat{L}\Big)$$

$$\log p(\vec{z}(t_1)) = \log p(\vec{z}(t_0)) - \underline{\mathbb{E}_{p(\vec{\epsilon})}\left[\int_{t_0}^{t_1} \vec{\epsilon}^{\,T} \frac{\partial f}{\partial \vec{z}} \vec{\epsilon}\, dt\right]}$$

Experiments

Density estimation on toy data

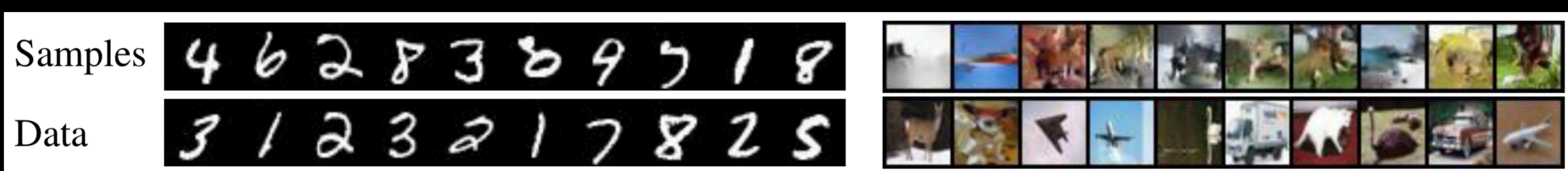


Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Density estimation on real data

	POWER	GAS	HEPMASS	MINIBOONE	BSDS300	MNIST	CIFAR10
Real NVP	-0.17	-8.33	18.71	13.55	-153.28	1.06*	3.49*
Glow	-0.17	-8.15	18.92	11.35	-155.07	1.05*	**3.35***
FFJORD	**-0.46**	**-8.59**	**14.92**	**10.43**	**-157.40**	**0.99*** ($1.05^{\dagger}$)	3.40*
MADE	3.08	-3.56	20.98	15.59	-148.85	2.04	5.67
MAF	-0.24	-10.08	17.70	11.75	-155.69	1.89	4.31
TAN	-0.48	-11.19	15.12	11.01	-157.03	-	-
MAF-DDSF	-0.62	-11.96	15.09	8.86	-157.73	-	-

Table 2: Negative log-likehood on test data for density estimation models; **lower is better**. In nats for tabular data and bits/dim for MNIST and CIFAR10. *Results use multi-scale convolutional architectures. †Results use a single flow with a convolutional encoder-decoder architecture.



Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

VAE with FFJORD

	MNIST	Omniglot	Frey Faces	Caltech Silhouettes
No Flow	$86.55 \pm .06$	$104.28 \pm .39$	$4.53 \pm .02$	$110.80 \pm .46$
Planar	$86.06 \pm .31$	$102.65 \pm .42$	$4.40 \pm .06$	$109.66 \pm .42$
IAF	$84.20 \pm .17$	$102.41 \pm .04$	$4.47 \pm .05$	$111.58 \pm .38$
Sylvester	$83.32 \pm .06$	$99.00 \pm .04$	$4.45 \pm .04$	$104.62 \pm .29$
FFJORD	$\mathbf{82.82} \pm \mathbf{.01}$	$\mathbf{98.33} \pm \mathbf{.09}$	$\mathbf{4.39} \pm \mathbf{.01}$	$\mathbf{104.03} \pm \mathbf{.43}$

Table 3: Negative ELBO on test data for VAE models; **lower is better**. In nats for all datasets except Frey Faces which is presented in bits per dimension. Mean/stdev are estimated over 3 runs.

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

	Method	Train on data	One-pass Sampling	Exact/Unbiased Log-likelihood	Free-form Jacobian
	Variational Autoencoders	✓	✓	✗	✓
	Generative Adversarial Nets	✓	✓	✗	✓
	Likelihood-based Autoregressive	✓	✗	✓	✗
Change of Variables	Normalizing Flows	✗	✓	✓	✗
	Reverse-NF, MAF, TAN	✓	✗	✓	✗
	NICE, Real NVP, Glow, Planar CNF	✓	✓	✓	✗
	FFJORD	✓	✓	✓	✓

Table 1: A comparison of recent generative modeling approaches.

Grathwohl, Will, et al. "Ffjord: Free-form continuous dynamics for scalable reversible generative models."

Summary

- Normalizing flows help us transform density
- Continuous NF do it better than NF
- FFJORD do it better than CNF

Summary

- Normalizing flows help us transform density
- Continuous NF do it better than NF
- FFJORD do it better than CNF