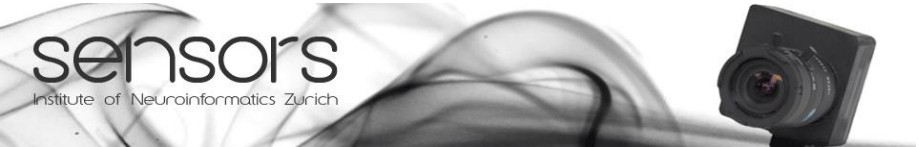
DeltaDEQ: Exploiting Heterogeneous Convergence for Accelerating Deep Equilibrium Iterations

Zuowen Wang, Longbiao Cheng, Pehuen Moure, Niklas Hahn, Shih-Chii Liu

Institute of Neuroinformatics
University of Zurich and ETH Zurich



Implicit Neural Networks

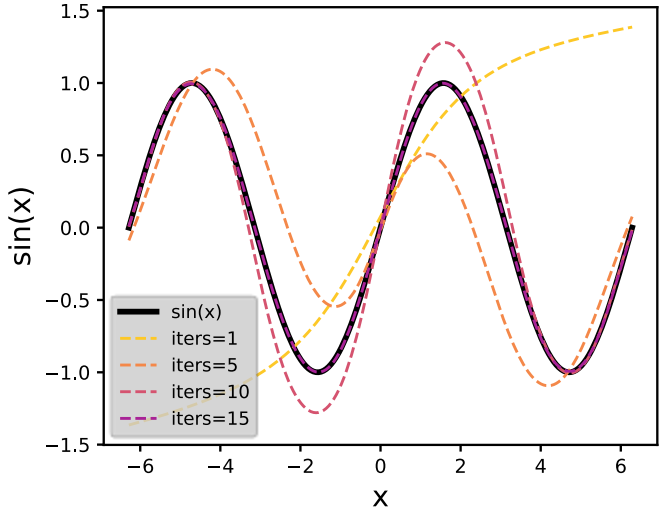
- Implicit neural networks, have gained much attention:
  - Computer vision, language modeling, audio processing, etc.
  - Representative work: Deep Equilibrium Models (DEQ)
  - $z^* = f_\theta(z^*, x)$
  - This fixed-point equation can be solved by $z^{i+1} = f_\theta(z^i, x)$

- Problem: Inference requires expensive fixed-point equation solving!
  - Each fixed-point iteration = one forward-pass

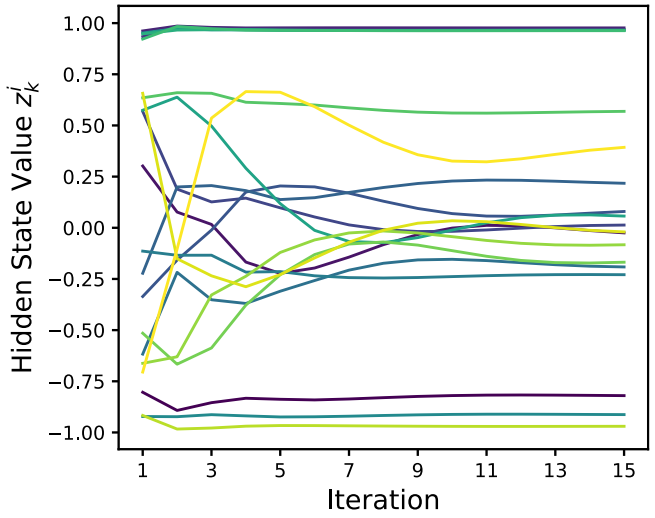
[1] Shaojie Bai et al. “Deep Equilibrium Models” 33rd Conference on Neural Information Processing Systems 2019

Heterogeneous Convergence

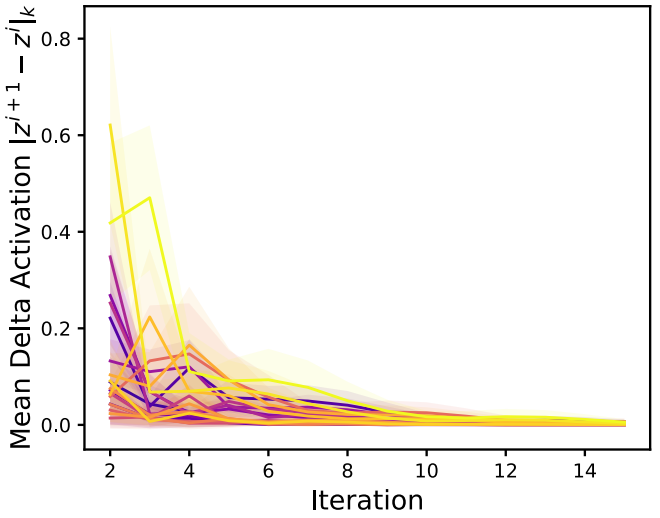
- Observe elementwise activation evolution in DEQ
  - Different dimensions of activations converge with different speed
  - Considering a sin(x) function fitting example with $z^* = \tanh(W_z \cdot z^* + W_x \cdot x + b)$



(a)



(b)



(c)

Delta Deep Equilibrium Layer (DeltaDEQ)

- 'Delta-ize' the linear operations in a DEQ layer
  - RNN-type example:

$$z_t^{i+1} = \sigma(W_z \cdot z_t^i + W_x \cdot x_t) = \sigma(W_z \cdot (z_t^i - z_t^{i-1}) + W_z \cdot z_t^{i-1} + W_x \cdot x_t)$$

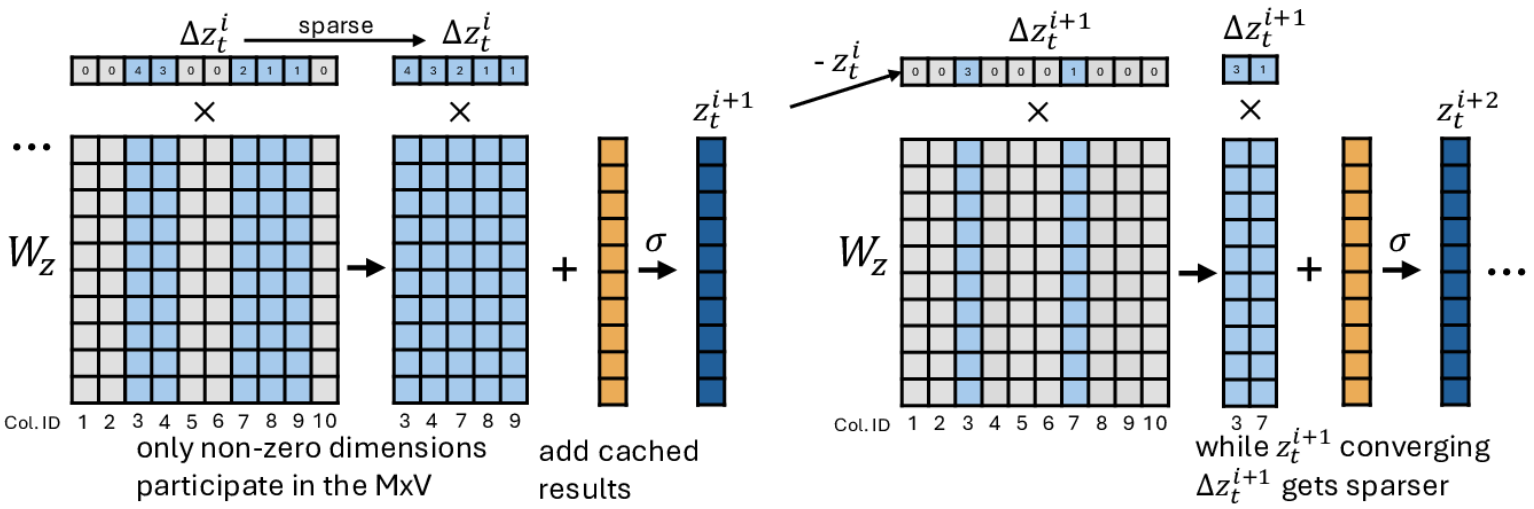
  - Apply an element-wise delta threshold on the difference

$$\Delta z_t^i := \begin{cases} z_t^i - z_t^{i-1} & \text{if } |z_t^i - z_t^{i-1}| \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (\textbf{Delta Rule})$$

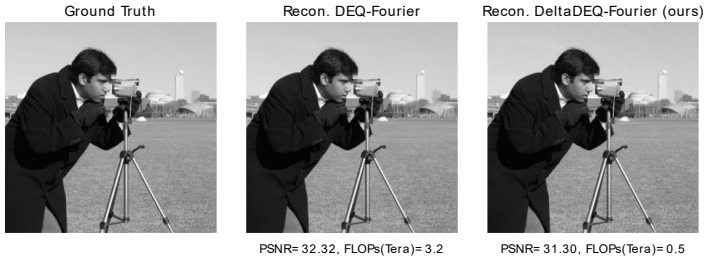
$$z_t^{i+1} \approx \sigma(\underbrace{W_z \cdot \Delta z_t^i}_{\text{sparse}} + \underbrace{W_z \cdot z_t^{i-1}}_{\text{cached in } C_z} + \underbrace{W_x \cdot x_t}_{\text{cached in } C_x}).$$

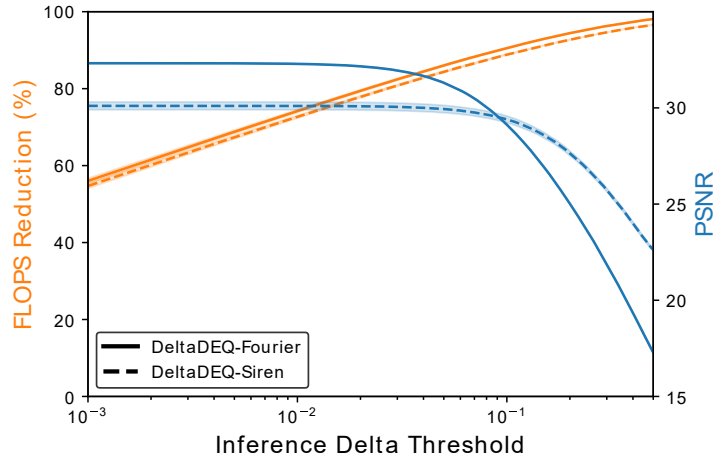
$$C_z \leftarrow W_z \cdot \Delta z_t^i + C_z$$

Saving on computation

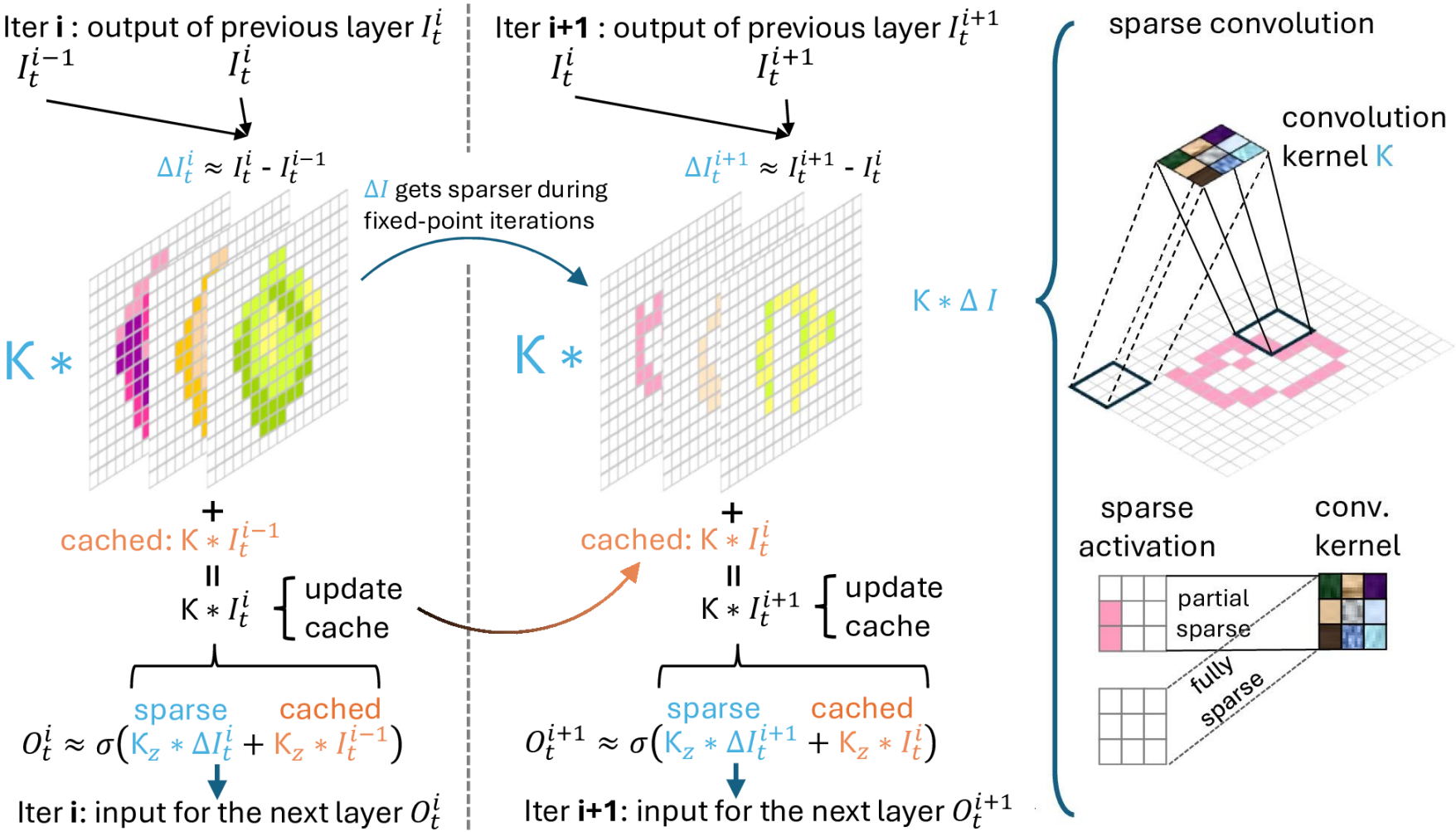


$$z_t^{i+1} \approx \sigma(\underbrace{W_z \cdot \Delta z_t^i}_{\text{sparse}} + \underbrace{W_z \cdot z_t^{i-1}}_{\text{cached in } C_z} + \underbrace{W_x \cdot x_t}_{\text{cached in } C_x}).$$





DeltaDEQ for Convolutional Layers



Thanks!

- For more details, please refer to our paper.