Ode to an ODE

Krzysztof Choromanski, Jared Quincy Davis, Valerii Likhosherstov, Xingyou Song, Jean-Jacques Slotine, Jacob Varley, Honglak Lee, Adrian Weller, Vikas Sindhwani
Neural ODEs:

- Continuous variants of standard ResNet networks:

\[
\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}_t, t, \theta) \quad \mathbf{x}_t = \mathbf{x}_{t_0} + \int_{t_0}^{t} f(\mathbf{x}_s, s, \theta) ds \quad (1)
\]

- Emulate deep discrete neural networks with **compact** number of parameters.
- Parameters of the Neural ODEs encapsulated in the mapping \(\theta(t)\).

**How to design it?**

- As every deep neural network system, suffer from exploding/vanishing gradients which makes training challenging. Can we robustify Neural ODEs?
Ode to an ODE System:

- **IDEA:** Design $\theta(t)$, so that when integrated, Neural ODE emulates deep ResNet with **orthogonal** connection matrices.

- This leads to the matrix-flow on the **orthogonal group** and effectively: to a **nested system of flows**, where the orthogonal flow encoding $\theta(t)$ determines main flow. How to design learnable orthogonal flows and why are they good?
Orthogonal Flows:

\begin{equation}
\begin{cases}
\dot{x}_t = f(W_t x_t) \\
\dot{W}_t = W_t b_\psi(t, W_t)
\end{cases}
\end{equation}

mapping to skew-symmetric matrices

- $b_\psi$ can be modeled by a neural network producing skew-symmetric matrices or via parameterized isospectral flows (e.g. double-bracket flows)

\underline{Lemma 4.1} (ODEtoODES for gradient stabilization). Consider a Neural ODE on time interval $[0, T]$ and given by Formula 2. Let $\mathcal{L} = \mathcal{L}(x_T)$ be a differentiable loss function. The following holds for any $t \in [0, 1]$, where $e = 2.71828...$ is Euler constant:

$$
\frac{1}{e} \left\| \frac{\partial \mathcal{L}}{\partial x_T} \right\|_2 \leq \left\| \frac{\partial \mathcal{L}}{\partial x_t} \right\|_2 \leq e \left\| \frac{\partial \mathcal{L}}{\partial x_T} \right\|_2.
$$
Ode to an ODE System in Action:

**RL:** comparison with Deep(Res)Nets, NANODE, Base ODEs, and ANODEV2-hypernets

**Supervised:** MNIST-Corrupted

<table>
<thead>
<tr>
<th>Models</th>
<th>Dense-1</th>
<th>Dense-10</th>
<th>NODE</th>
<th>NANODE-1</th>
<th>NANODE-10</th>
<th>HyperNet-1</th>
<th>HyperNet-10</th>
<th>ODEtoODE-1</th>
<th>ODEtoODE-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dotted Lines</td>
<td>92.99</td>
<td>88.22</td>
<td>91.54</td>
<td>92.42</td>
<td>92.74</td>
<td>91.88</td>
<td>91.9</td>
<td>95.42</td>
<td>95.22</td>
</tr>
<tr>
<td>Spatter</td>
<td>93.54</td>
<td>89.52</td>
<td>92.49</td>
<td>93.13</td>
<td>93.15</td>
<td>93.19</td>
<td>93.28</td>
<td>94.9</td>
<td>94.9</td>
</tr>
<tr>
<td>Stripe</td>
<td>30.55</td>
<td>20.57</td>
<td>36.76</td>
<td>16.4</td>
<td>21.37</td>
<td>19.71</td>
<td>18.69</td>
<td>44.51</td>
<td>44.37</td>
</tr>
<tr>
<td>Translate</td>
<td>25.8</td>
<td>24.82</td>
<td>27.09</td>
<td>28.97</td>
<td>29.31</td>
<td>29.42</td>
<td>29.3</td>
<td>26.82</td>
<td>26.63</td>
</tr>
<tr>
<td>Rotate</td>
<td>82.8</td>
<td>80.38</td>
<td>82.76</td>
<td>83.19</td>
<td>83.65</td>
<td>83.5</td>
<td>83.56</td>
<td>83.9</td>
<td>84.1</td>
</tr>
<tr>
<td>Scale</td>
<td>58.68</td>
<td>62.73</td>
<td>62.05</td>
<td>66.43</td>
<td>66.63</td>
<td>67.84</td>
<td>68.11</td>
<td>66.68</td>
<td>66.76</td>
</tr>
<tr>
<td>Shear</td>
<td>91.73</td>
<td>89.52</td>
<td>91.82</td>
<td>92.18</td>
<td>93.71</td>
<td>92.33</td>
<td>92.48</td>
<td>93.37</td>
<td>92.93</td>
</tr>
<tr>
<td>Motion Blur</td>
<td>78.16</td>
<td>67.25</td>
<td>75.18</td>
<td>76.53</td>
<td>79.22</td>
<td>78.82</td>
<td>78.33</td>
<td>78.63</td>
<td>77.58</td>
</tr>
<tr>
<td>Glass Blur</td>
<td>91.62</td>
<td>84.89</td>
<td>87.94</td>
<td>90.18</td>
<td>91.07</td>
<td>91.3</td>
<td>91.17</td>
<td>93.91</td>
<td>93.28</td>
</tr>
<tr>
<td>Shot Noise</td>
<td>96.16</td>
<td>91.63</td>
<td>94.24</td>
<td>94.97</td>
<td>94.73</td>
<td>94.76</td>
<td>94.81</td>
<td>96.91</td>
<td>96.71</td>
</tr>
<tr>
<td>Identity</td>
<td>97.55</td>
<td>95.73</td>
<td>97.61</td>
<td>97.65</td>
<td>97.69</td>
<td>97.72</td>
<td>97.54</td>
<td>97.94</td>
<td>97.88</td>
</tr>
</tbody>
</table>
Thank you for your Attention!