Predictions Based on Pixel Data

TES Conference on Mathematical Optimization for Machine Learning

Davide Murari
davide.murari@ntnu.no
In collaboration with Elena Celledoni, James Jackaman, Brynjulf Owren

NTNU
Norwegian University of Science and Technology

Definition of the problem

We have some data

$$\{(U_i^0, ..., U_i^M)\}_{i=1}^N$$

We know there is a relation of the type

$$U_i^{j+1} = \Psi(U_i^j) + \delta_i^{j+1} \in \mathbb{R}^{d\times d}$$

We then want to approximate such map

$$\mathcal{N}_\theta \approx \Psi$$

Relevant questions

Consider the parametric space of functions

$$\mathcal{F} = \{\mathcal{N}_\theta : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : \ \theta \in \Theta\}$$

1. How well can we approximate the desired map with functions in this space?

2. Can we find the best approximator given the data we have available?

Main questions

Consider the parametric space of functions

$$\mathcal{F} = \{\mathcal{N}_\theta : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : \ \theta \in \Theta\}$$

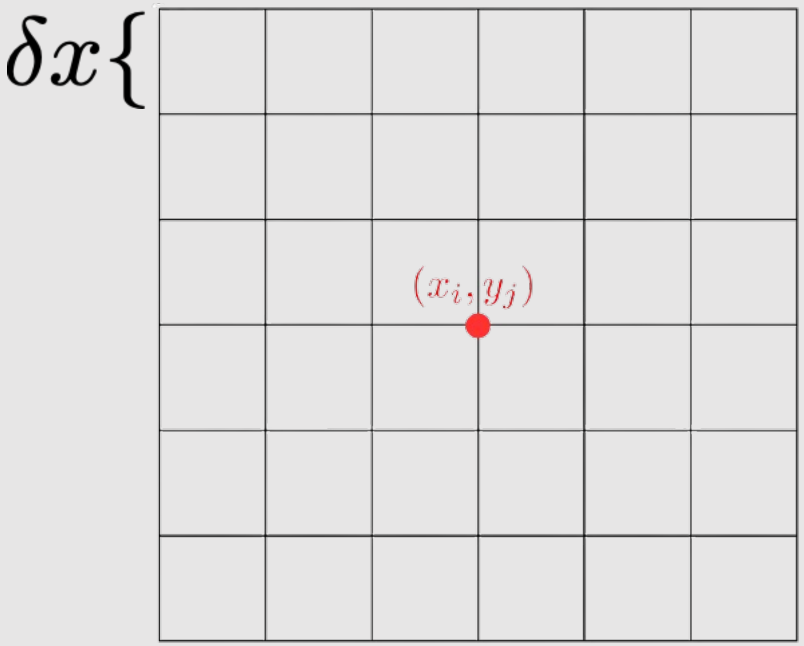
1. How well can we approximate the desired map with functions in this space?

2. Can we find the best approximator given the data we have available?

Particular case of interest

$$u = u(t, x, y) \in \mathbb{R},\ \ \beta_i \in \mathbb{R},\ \ \alpha_{ij} \in \mathbb{N}^2$$

$$\partial_t u = \mathcal{L}u + \sum_{i=1}^{n} \beta_i (\partial_{\alpha_{i1}} u)(\partial_{\alpha_{i2}} u)$$



$$0 = t_0 < t_1 < \ldots < t_M$$

$$(U^j)_{rs} = u(t_j, x_r, y_s)$$

$$x_r = r\delta x,\ \ y_s = s\delta x.$$

Method of lines

Time dependent PDE

$$\partial_t u = \mathcal{L}u + \sum_{i=1}^{n} \beta_i (\partial_{\alpha_{i1}} u)(\partial_{\alpha_{i2}} u)$$

Method of lines

Time dependent PDE

$$\partial_t u = \mathcal{L}u + \sum_{i=1}^{n} \beta_i (\partial_{\alpha_{i1}} u)(\partial_{\alpha_{i2}} u)$$

Semi-discretization in space with finite differences

$$\dot{U}(t) = L * U(t) + \sum_{i=1}^{n} \beta_i (D_{i1} * U(t)) \odot (D_{i2} * U(t)) := F(U(t))$$

Method of lines

Time dependent PDE

$$\partial_t u = \mathcal{L}u + \sum_{i=1}^{n} \beta_i (\partial_{\alpha_{i1}} u)(\partial_{\alpha_{i2}} u)$$

Semi-discretization in space with finite differences

$$\dot{U}(t) = L * U(t) + \sum_{i=1}^{n} \beta_i (D_{i1} * U(t)) \odot (D_{i2} * U(t)) := F(U(t))$$

Numerical discretization of the space-time dynamics

$$U_i^{j+1} = U_i^j + \delta t F(U_i^j) =: \Psi_F^{\delta t}(U_i^j)$$

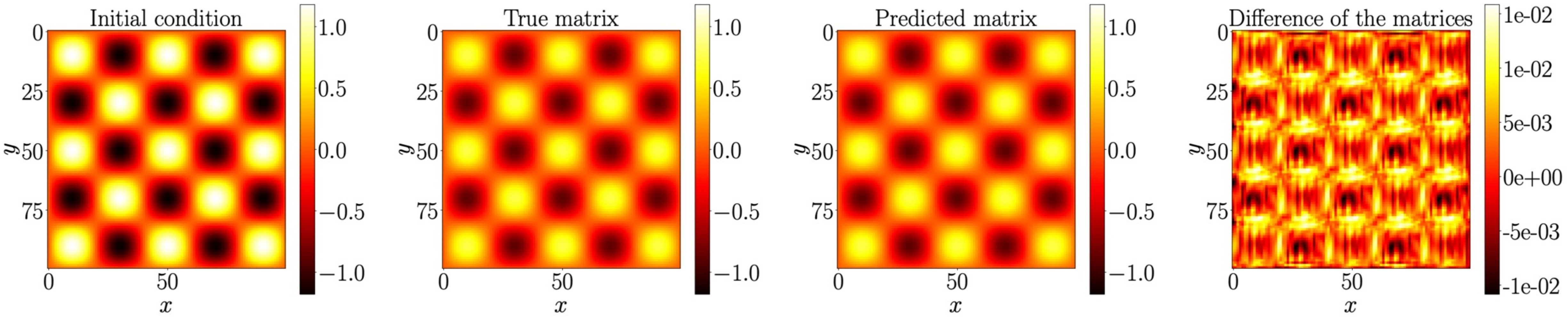
Learning the solution map

Data:

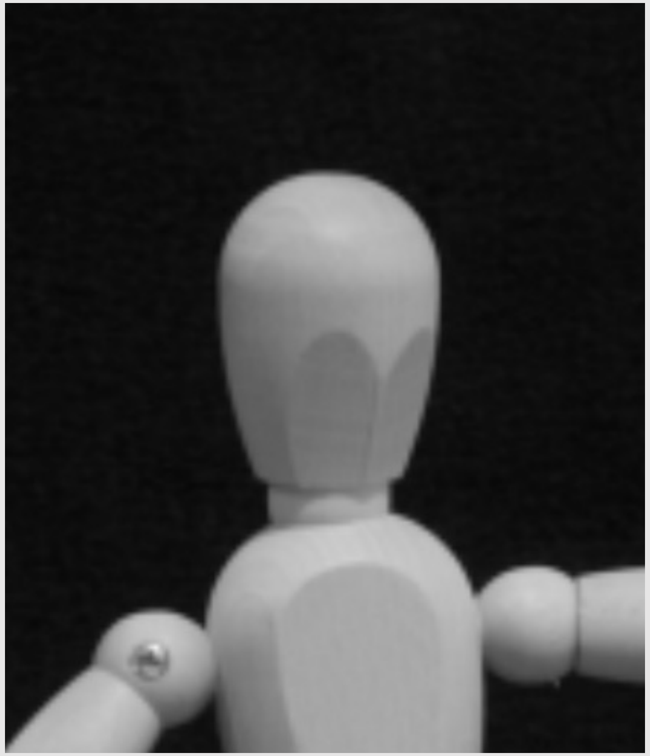
$$U_i^{j+1} = \Phi_F^{\delta t}(U_i^j) + \mathcal{O}(\delta x^2)$$

Fisher's Equation:

$$\partial_t u = \alpha \Delta u + u(1-u)$$



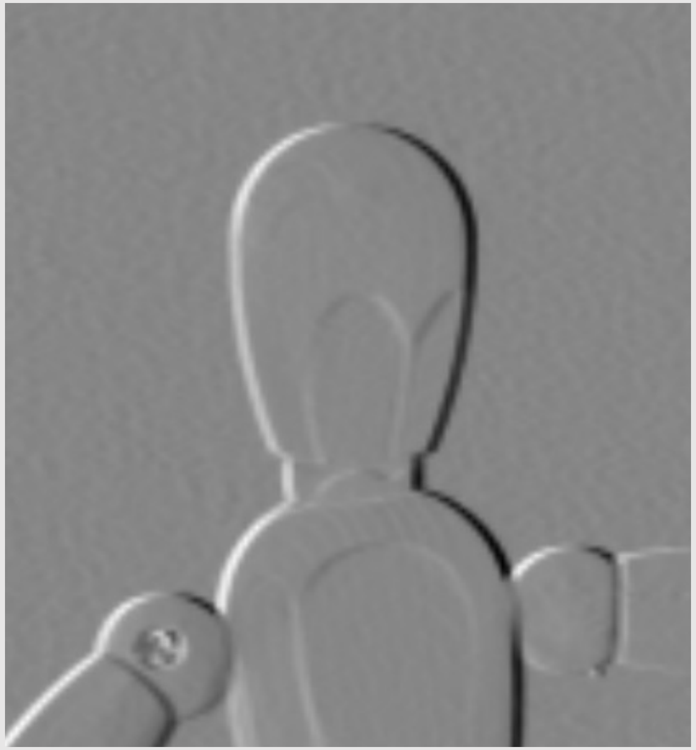
Finite differences as convolution



$$* \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} =$$

$$\wr\wr$$

$$-\partial_x$$



Space of functions

$$\begin{cases} \dot{V}(t) = F_\theta(V(t)) \in \mathbb{R}^{d\times d}, & \theta \in \Theta \\ V(0) = V_0 \end{cases}$$

$$\Phi^{\delta t}_{F_\theta}(V_0) = V(\delta t) = \Psi^{\delta t}_{F_\theta}(V_0) + \mathcal{O}(\delta t^{r+1})$$

$$\mathcal{F} = \{\Psi^{\delta t}_{F_\theta} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d} : \theta \in \Theta,\ F_\theta \text{ is a CNN}\}$$

Example : Explicit Euler gives a ResNet

$$\Psi^{\delta t}_{F_\theta}(V) = V + \delta t F_\theta(V)$$

Error splitting

$$\left\|U^{j+1} - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\|$$

$$= \left\|\mathcal{O}\left(\delta x^2\right) + \Phi_F^{\delta t}\left(U^j\right) - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\|$$

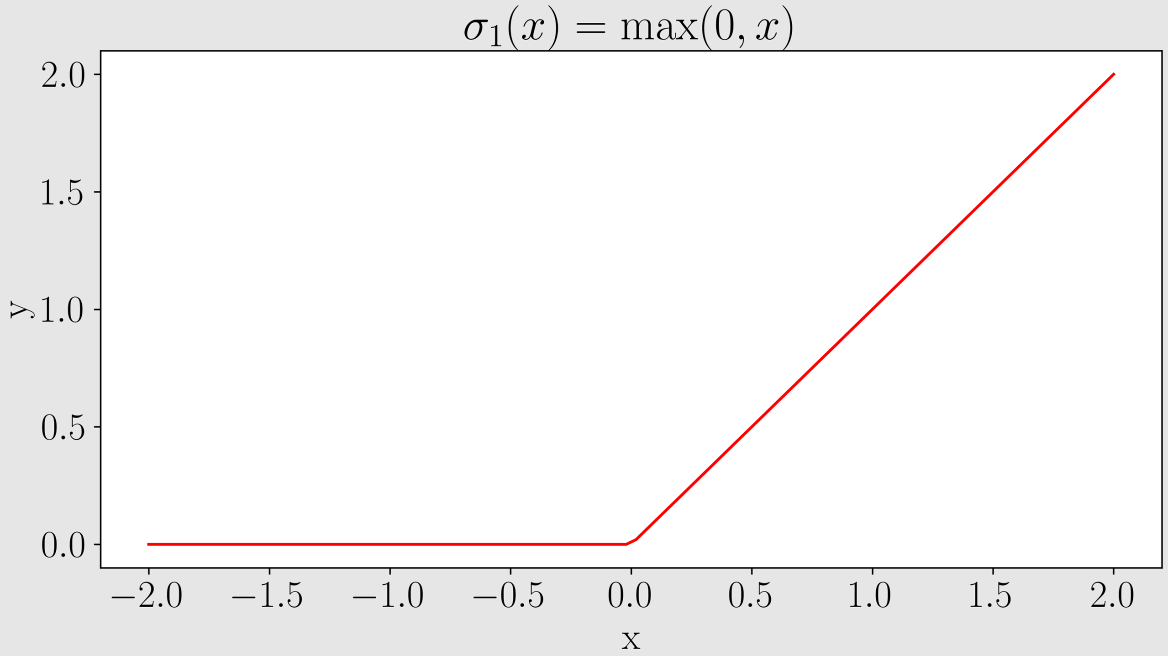
$$\leq \mathcal{O}\left(\delta x^2\right) + \left\|\Phi_F^{\delta t}\left(U^j\right) - \Psi_F^{\delta t}\left(U^j\right) + \Psi_F^{\delta t}\left(U^j\right) - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\|$$

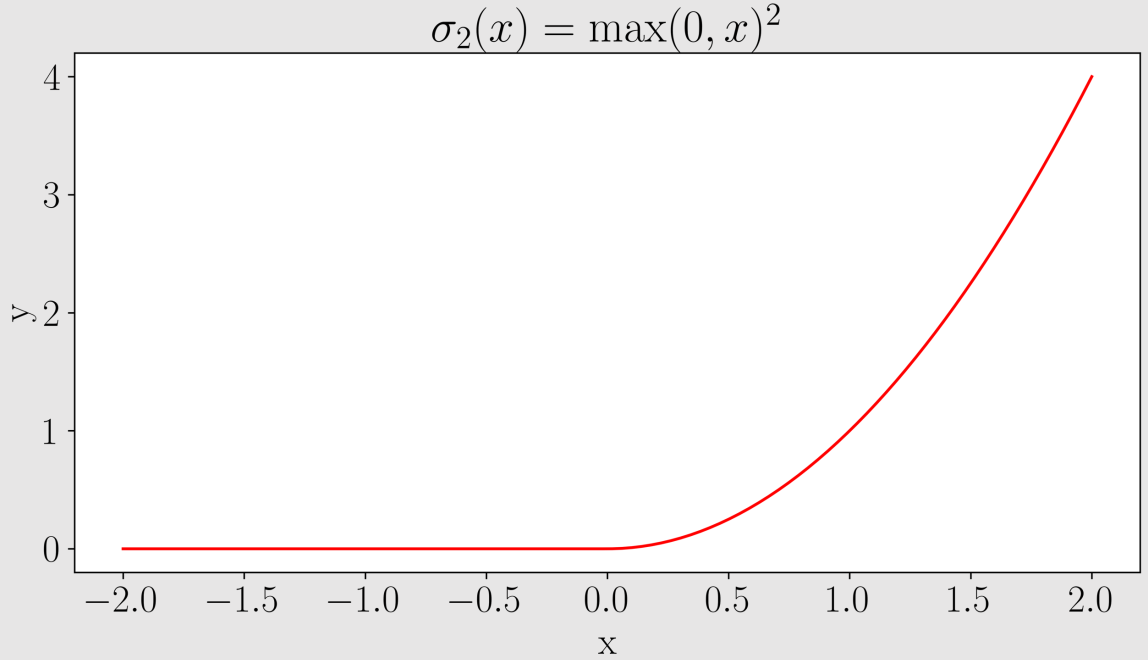
$$\leq \underbrace{\mathcal{O}\left(\delta x^2\right)}_{\text{spatial error}} + \underbrace{\left\|\Phi_F^{\delta t}\left(U^j\right) - \Psi_F^{\delta t}\left(U^j\right)\right\|}_{\text{classical error estimate}} + \underbrace{\left\|\Psi_F^{\delta t}\left(U^j\right) - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\|}_{\text{neural network approximation}}$$

$$\leq \mathcal{O}\left(\delta x^2\right) + \mathcal{O}\left(\delta t^{r+1}\right) + \left\|\Phi_F^{\delta t}\left(U^j\right) - \Phi_{F_\theta}^{\delta t}\left(U^j\right)\right\|$$

$$\leq \mathcal{O}\left(\delta x^2\right) + \mathcal{O}\left(\delta t^{r+1}\right) + \delta t \exp(\operatorname{Lip}(F)\delta t) \sup_{\|U\|\leq c} \left\|F_\theta(U) - F(U)\right\|$$

Useful activation functions





$$a = \sigma_1(a) - \sigma_1(-a)$$

$$\begin{aligned} 2ab &= (a+b)^2 - a^2 - b^2 \\ &= [\sigma_2(a+b) + \sigma_2(-(a+b)] \\ &+ [\sigma_2(a) + \sigma_2(-a)] + [\sigma_2(b) + \sigma_2(-b)] \end{aligned}$$

Simplified setting

Example:

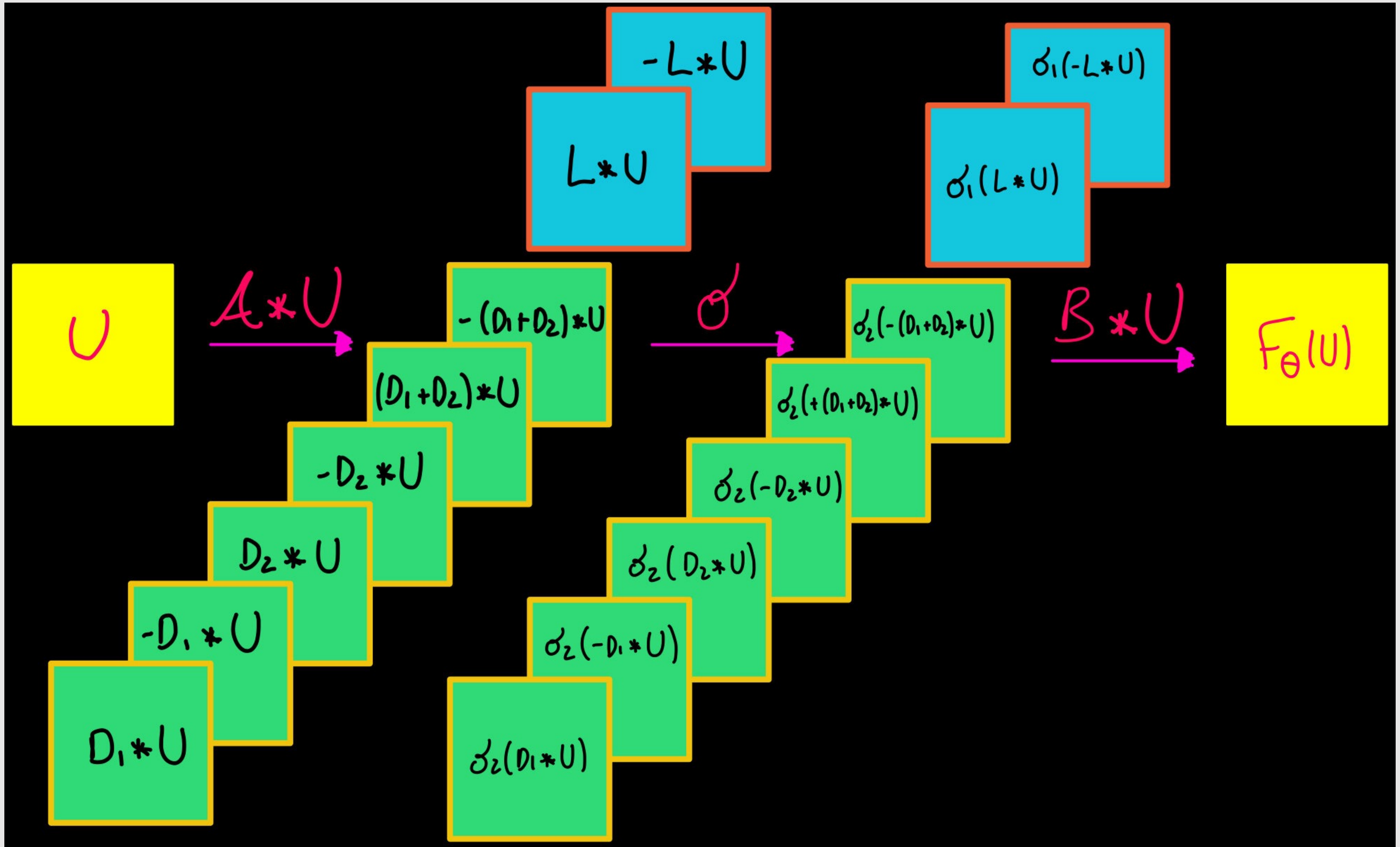
$$\partial_t u = \partial_x u + u\partial_y u$$

Semidiscretization:

$$F(U) = L * U + (D_1 * U) \odot (D_2 * U)$$

Representation with CNN

$$F(U) = F_\theta(U) := \mathcal{B} * \sigma(\mathcal{A} * U)$$



Approximation theorem

$$\left\|U^{j+1} - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\|$$
$$\leq \mathcal{O}\left(\delta x^2\right) + \mathcal{O}\left(\delta t^{r+1}\right) + \delta t \exp(\mathrm{Lip}(F)\delta t) \sup_{\|U\|\leq c} \|F_\theta(U) - F(U)\|$$

Approximation theorem

$$\left\|U^{j+1} - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\|$$
$$\leq \mathcal{O}\left(\delta x^2\right) + \mathcal{O}\left(\delta t^{r+1}\right) + \delta t \exp(\operatorname{Lip}(F)\delta t) \sup_{\|U\|\leq c} \|F_\theta(U) - F(U)\|$$

There is a 2-layer CNN

$$F_\theta$$

making this quantity vanish

Approximation theorem

$$\left\|U^{j+1} - \Psi_{F_\theta}^{\delta t}\left(U^j\right)\right\| \leq \mathcal{O}\left(\delta x^2\right) + \mathcal{O}\left(\delta t^{r+1}\right)$$

$$\Downarrow$$

$$\min_{\Psi_{F_\theta}^{\delta t} \in \mathcal{F}} \|\Psi_{F_\theta}^{\delta t} - \Phi_F^{\delta t}\| \leq \mathcal{O}(\delta x^2) + \mathcal{O}(\delta t^{r+1})$$

Conclusion

1. How well can we approximate the desired map with functions in this space?

If there is no noise, at least as well as with the method of lines and the correct right hand side

2. Can we find the best approximator given the data we have available?

Thank you for the attention

Celledoni, E., Jackaman, J., M., D., & Owren, B., preprint (2023)
Predictions Based on Pixel Data: Insights from PDEs and Finite Differences.