SAND2019-15055C

aboratories

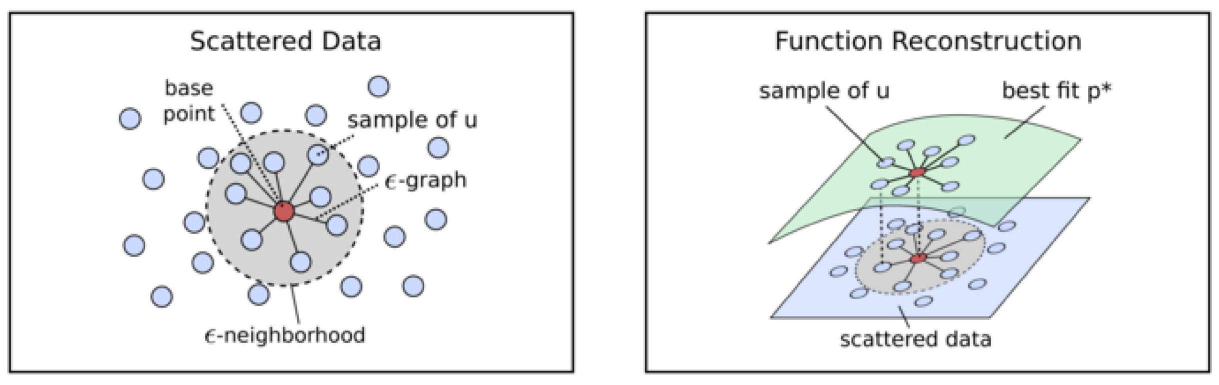
entific Machine Learning Methods for Unstructured Data

a National Laboratories), B. Gross, P. Atzberger (UC Santa Barbara)

Introduction

- Scientific machine learning requires handling data that is unstructured, small, and heterogeneous
- Seek a generalization of convolutional network to exploit weight-sharing with requiring grid
- **Idea:** build a tool on top of a rigorous scattered data approximation theory to obtain a parameterized function estimator

Generalized Moving Least Squares



$$p^* = \underset{p \in \mathbb{P}}{\operatorname{argmin}} \sum_{j=1}^{N} (\lambda_j(u) - \lambda_j(p))^2 \, \omega(||x - x_j||)$$

$$p^* = \Phi(x)^\intercal \mathbf{a}(u)$$

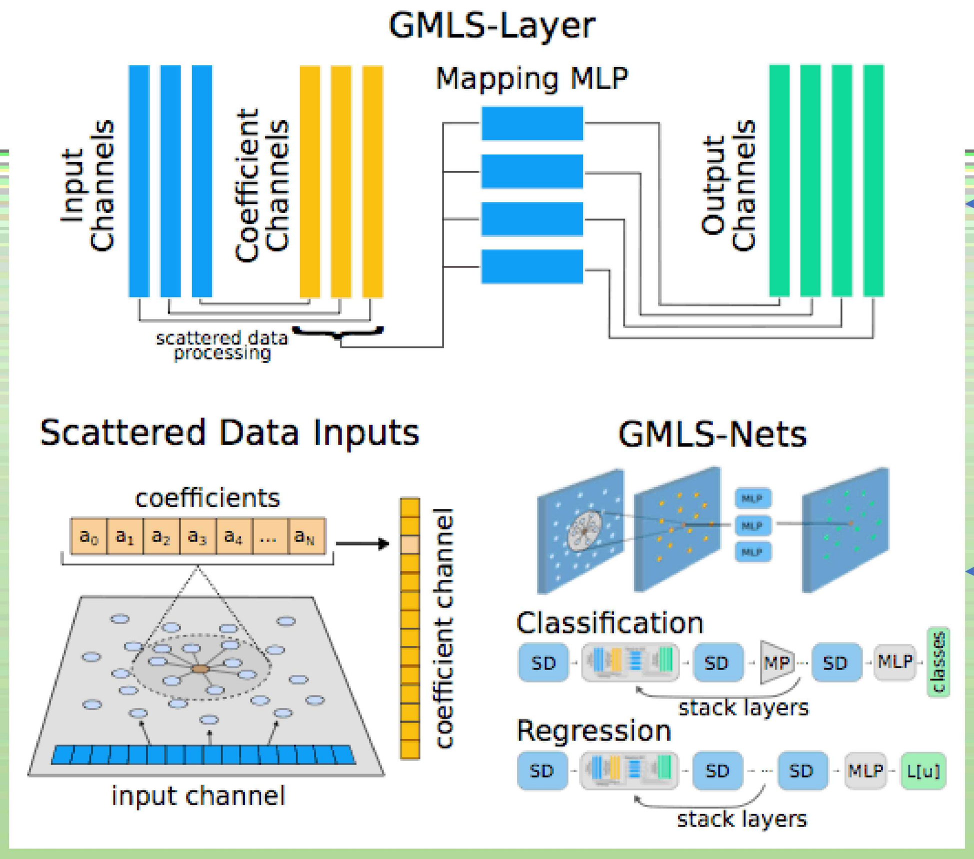
$$\tau^h_{\tilde{\mathbf{x}}}[u] = \tau_{\tilde{\mathbf{x}}}(\Phi)^\intercal \mathbf{a}(u)$$

How to estimate a functional from scattered data:

- Given samples $\{\lambda\}$, target τ, compact $\omega > 0$ and reconstruction space $\mathbb{P}$
- Solve local least squares problem to get optimal reconstruction
- Apply τ to reconstruction to get estimate

GMLS-Nets: Operator learning architecture

- Function estimation requires knowledge of how τ acts on data
- By parameterizing $\tau(\Phi)$ we learn how operator acts on basis
- Direct analogue to learning stencil in ConvNets, but no requirement on data layout or boundary treatment
- Exploit regularity in SciML applications - for smooth solutions of PDE learning action of operator on polynomial basis is much lower dimensional than learning action on whole function space



Parameterize action of operator on basis with either a linear layer or deep MLPs

Stackable layers exploiting weight-sharing, may be used same as ConvNets but on scientific data

Software available at https://github.com/rgp62/gmls-nets (Tensorflow) and https://github.com/atzberg/gmls-nets (PyTorch)

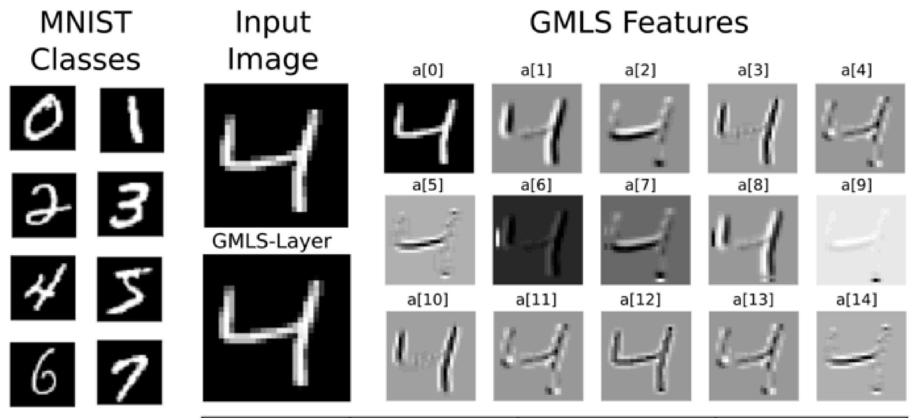
```
model = tf.keras.Sequential()
model.add(tf.keras.layers.Reshape((n**2,1),input_shape=(n,n)))

model.add(gnets.MFConvLayer(x1,x2,fP,gnets.
              weightfuncs.sixth,eps1,chans1,activation='elu'))
model.add(tf.keras.layers.BatchNormalization(-1))

model.add(gnets.MFPoolLayer(x2,x3,eps2,tf.reduce_max))

model.add(tf.keras.layers.Flatten())
model.add(tf.keras.layers.Dense(100, activation='elu'))
model.add(tf.keras.layers.Dense(10, activation='softmax'))
```

Strided Convolution → Max Pool → Dense (elu) → Dense (softmax)

Applications

MNIST: ConvNet Comparison

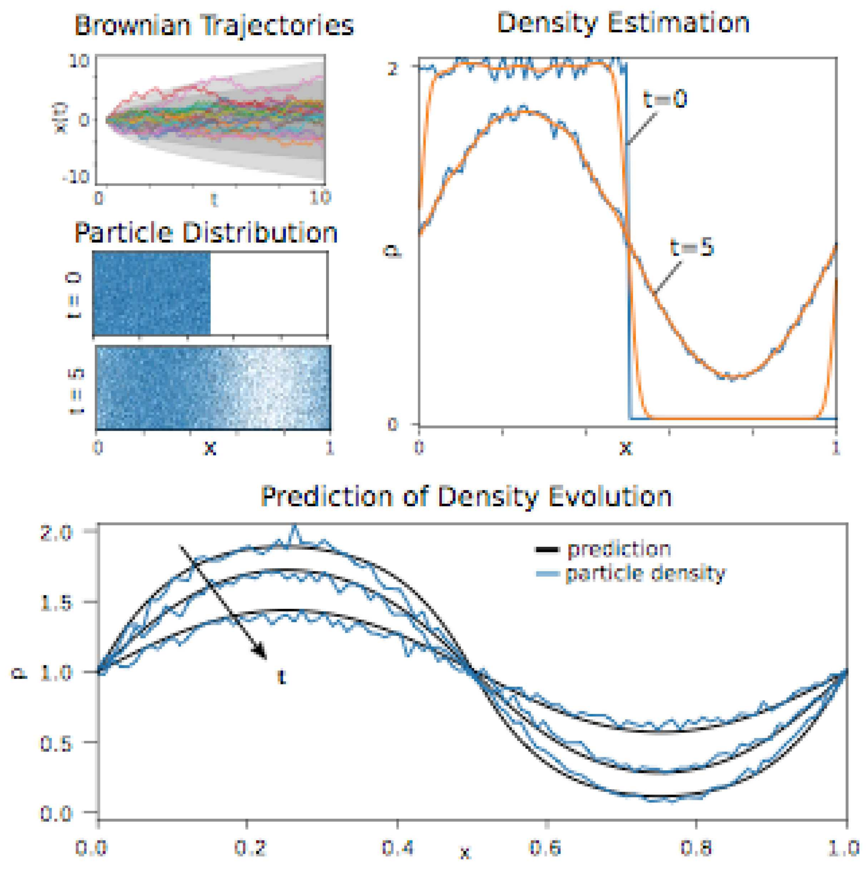
- On regular grid, comparable performance to ConvNets on classification problems
  - Channels similarly extract features



Case	Conv-2L	Hybrid-2L	GMLS-2L
MNIST	98.52%	98.41%	96.87%

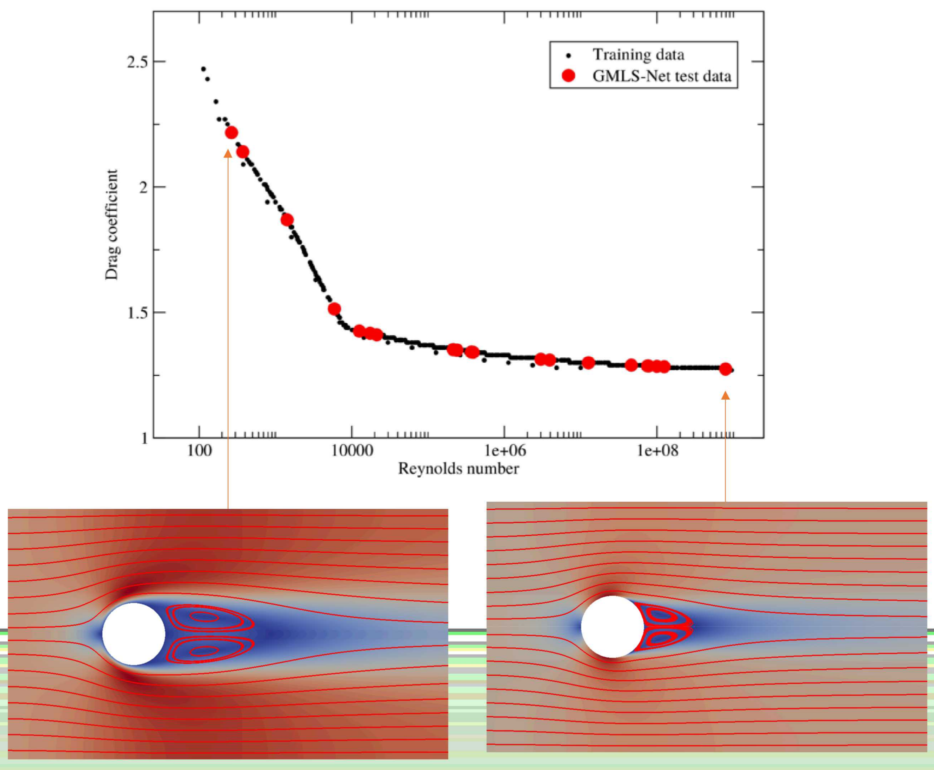
Data-driven model extraction

- Want to extract fast PDE models from high fidelity molecular simulations
  - Run a **single** molecular simulation of Brownian motion, histogram and filter particle positions to extract training data for evolution of density over a few timesteps
  - From one time window, obtain massive number of observations of how density evolves
  - Test by comparing short term operator against long term dynamics



Engineering QOI from incomplete simulation data

- Generate training data by running **unstructured** finite volume code and computing drag from simulated **velocity and pressure** field at cell centers
- Train GMLS-Net **using only velocity fields** as input and drag as output, with **no knowledge of viscosity or Reynolds number**
- Excellent agreement, despite fact that only partial data given
- In engineering practice, similar correlations may be obtained by dimensional analysis and correlating length of recirculation zone with Reynolds number and drag coefficient
- GMLS-Nets find similar connection between flow structure and drag without intervention
- Applicable to e.g. process velocimetry data



- **GMLS-Nets extend capabilities of ConvNets to scattered data representative of SciML**
- **By exploiting regularity, excellent agreement for science and engineering applications**
- **Future directions: physics-informed NNs, manifold learning**

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. ___________

Sandia National Laboratories