Deep learning and inverse problems for geosteering applications

Yuchen Jin, Qiuyang Shen, Xuqing Wu, Yueqin Huang and Jiefu Chen.
University of Houston, Cyentech Consulting LLC.

Geosteering Inverse Problem: Background

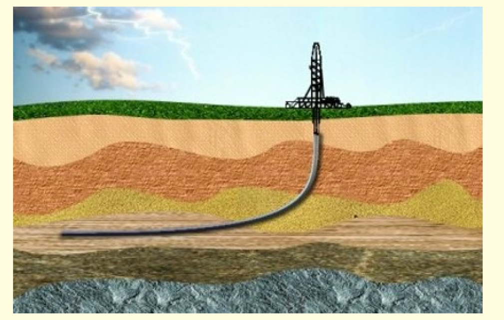


Figure: The schema of directional drilling. (www.amerexco.com)

- **Geosteering** is a key technique in directional drilling. It could be divided into two phases:
  - **Logging**: analyze the collected data.
  - **Drilling**: use the analysis results to adjust the well trajectory.

Geosteering Inverse Problem with a 3-layer Model

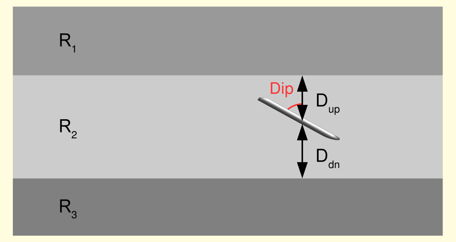


Figure: Directional drilling schema for an example of 3-layer model.

- The earth model are formulated by geophysical parameters.
  - R is resistivities.
  - $\mathrm{D_{up}}$ and $\mathrm{D_{dn}}$ are boundaries.
  - Dip is the dip angle (assumed to be fixed).
- The observed measurements are collected by the receiving antennas.

Common Solution to Geosteering Inverse Problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}) = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathcal{F}(\mathbf{x})\|_2^2 + \lambda\mathcal{R}(\mathbf{x}). \quad (1)$$

- The forward problem is of the form $\mathbf{y} = \mathcal{F}(\mathbf{x})$, where
  - $\mathbf{x} \in \mathbb{R}^5$ is earth model parameters including 3 resistivities and 2 boundaries.
  - $\mathbf{y} \in \mathbb{R}^{92}$ is observed measurements collected by logging tool.
  - $\mathcal{F} : \mathbb{R}^5 \mapsto \mathbb{R}^{92}$ is an EM forward model function defined by Maxwell Equations.
- Since $\mathcal{F}$ could produce synthetic measurements, (1) could be solved by minimizing the loss function $\mathcal{L}$, i.e. a synthetic error with a regularization term $\mathcal{R}(\mathbf{x})$.
- In deterministic methods [1, 2], $\mathbf{x}$ is optimized by calculating the gradient of $\mathcal{L}$.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 2\left(\mathbf{y} - \mathcal{F}(\mathbf{x})\right)\frac{\partial \mathcal{F}}{\partial \mathbf{x}} + \lambda\frac{\partial \mathcal{R}}{\partial \mathbf{x}}. \quad (2)$$

- (2) is used to calculate first-order gradient, where
  - $\partial\mathcal{F}/\partial\mathbf{x} \in \mathbb{R}^{92\times 5}$ is the *Jacobian* matrix.
  - $\partial\mathcal{R}/\partial\mathbf{x} \in \mathbb{R}^5$ is the gradient of regularization term.

Challenges for Logging

Table: Different logging methods.

	On ground	Underground
Data Amount	Inadequate	Adequate
Computation	Fast	Slow
Memory	Large	Small

- Two logging methods in practice:
- On ground method.
  - Process data on the ground.
  - Need to wait until the data gets transmitted back.
  - Data is not enough but hardware is powerful.
  - Use optimization method.
- Underground method.
  - Process data with devices underground.
  - Do not need to wait for data transmission.
  - All data is available but hardware is limited by underground environmental conditions.
  - Use lookup table as a surrogate.

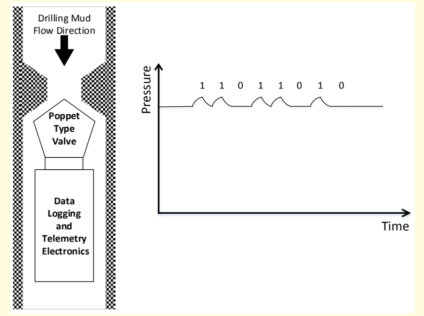


Figure: Positive purse pressure wave, used for data transmission, is with a very slow transmission rate.

Underground Method Surrogate: Lookup Table

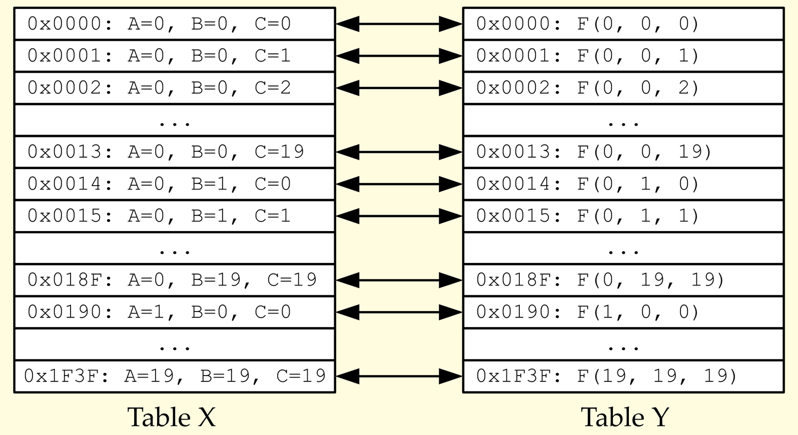


Figure: Lookup table method for fast estimation of the inversion.

- Construct table by pre-computing a lot of samples in the whole data space.
- Use the best-matched sample in the table to estimate a coarse solution.
- Drawbacks:
  - Large memory consumption and extremely coarse samples.
  - Both the computational cost and memory cost increase exponentially with the density of the table.

Our Proposed Surrogate: PhDNN

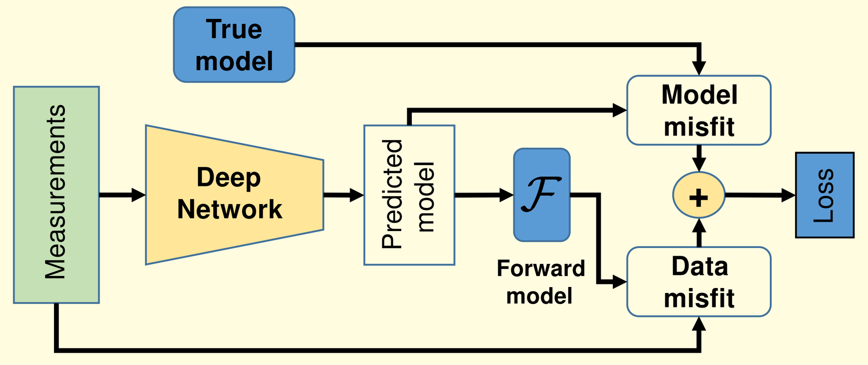


Figure: The physics-driven deep CNN (PhDNN) structure.

- The deep network is a 1D network which is adapted from VGG16 model. The model is trained by Adam optimizer [3].
- We have two contributions for this architecture.
  - Use 1D convolutional layers. Each layer composes of a convolution, an instance normalization [4] and a PReLu activation [5].
  - Use a physics-driven loss function:
    - A model misfit: $\|\mathbf{x} - \hat{\mathbf{x}}\|$.
    - A data misfit ($\|\mathbf{y} - \mathcal{F}(\hat{\mathbf{x}})\|$).

Computing Gradients for PhDNN – I

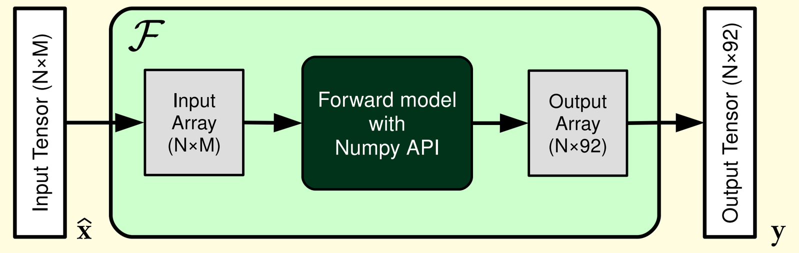


Figure: The implement of the forward model.

- The forward model function is highly nonlinear.
- In feed-forward process, the function could be viewed as a black box accepting earth model parameters ($\mathbf{x}$) and producing synthetic measurements ($\mathbf{y}$).

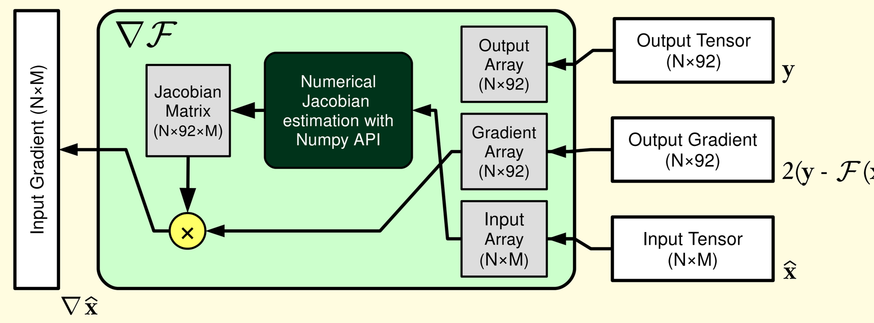


Figure: The implement of the back propagation of the forward model.

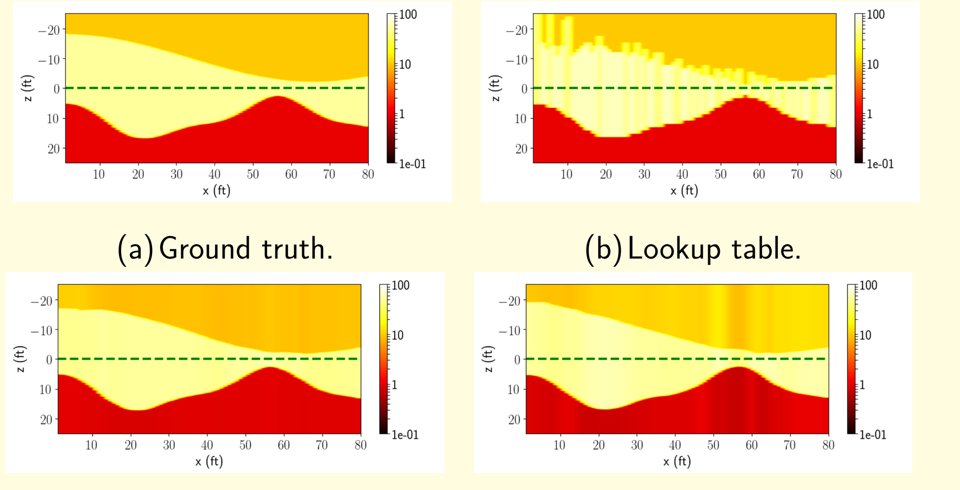
Computing Gradients for PhDNN – II

- In back-propagation process, The module accepts two inputs:
  - The current synthetic input of the forward model ($\hat{\mathbf{x}}$).
  - The gradient from the next layer ($2(\mathbf{y} - \mathcal{F}(\mathbf{x}))$).
- The gradient would be back-propagated to the previous layer in the deep network by using Jacobian matrix like (2).

Our Proposed Surrogate: Advantages – I

- The following performances are produced by a 2-core (i5) laptop without GPU.
- The network could be deployed for underground method.
  - The network is totally **feed-forward** and only requires **light computation** (about 0.3s for 80 points). The lookup table is slower (about 60s) while the optimization method is much slower (about 400s).
  - The network has a **small data size** (lower than 30MB) compared to a lookup table (about 1.6GB), which requires **lower memory consumption**.
- The network could make use of **all data** by taking advantage of underground method, while the optimization method could not.
- The network could get a far more **accurate prediction** compared to lookup table.
- The **computational cost** of the network **would not increase** with the data amount.

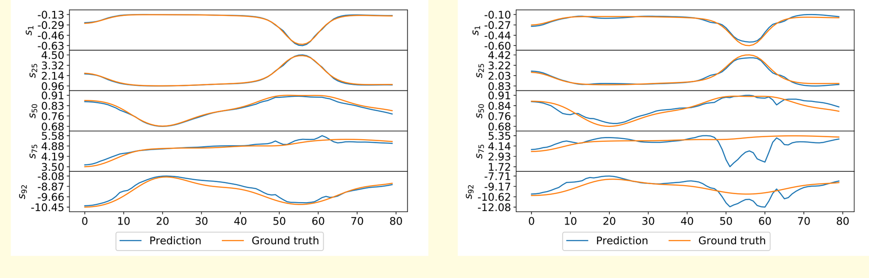
Results: Solution for Earth Model



(a) Ground truth. (b) Lookup table. (c) Proposed (PhDNN). (d) Data-driven Network.

Figure: The predicted earth model of an example, where the proposed network achieves better resistivity prediction compared to that of the conventional data-driven network.

Results: Synthetic Measurements



(a) Proposed (PhDNN). (b) Data-driven Network.

Figure: Some selected curves for the example, where the curves are picked from the synthetic measurements produced by the predicted earth model.

Results: Numerical Tests

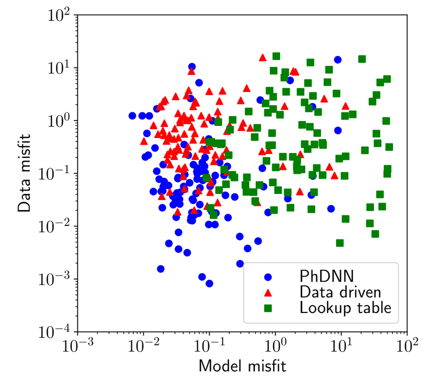


Figure: The numerical tests over compared methods.

- We generated 100 examples earth models like the shown one.
- $\boldsymbol{x}$ axis represents model misfit, and $\boldsymbol{y}$ axis represents data misfit.
- Compared to data-driven network, the proposed one could achieve the same model misfit but a better data misfit.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, and Office of Advanced Science Computing Research, under Award Numbers DE-SC0017033.

Reference

[1] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[2] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[3] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[4] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization. corr (2016)," *arXiv preprint arXiv:1607.08022*, 2016.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

SEG18 ANAHEIM, CA

UNIVERSITY of HOUSTON CULLEN COLLEGE of ENGINEERING