Overview:

- Machine learning can help improve results in seismic tomography
- Introduce machine learning and travel time tomography concepts
- Application of theory to data from the Long Beach array (Long Beach, CA, 2011)
Machine learning (ML)

Generally describes methods for
- Automatically detecting patterns in data
- Using patterns to predict future data or perform decision making
- ML relies on sufficient training data to obtain good performance

Methods and applications range from simple (and familiar) to complex

Linear regression

Segmentation of cells with convolutional neural network (CNN)

“U-net”
Machine learning for in geophysical tomography

Challenges include:

- Irregular sensor coverage
- Noise
- Little-no training data (ML approaches)

Other recent work:

- Kothari, Gupta, de Hoop, Dokmanic, "Random mesh projectors for inverse problems", ICLR 2019
Conventional least squares inversion

Inverse problem

Measurements → Model

"Measurement model"

\[ d = A m + n \]

Example model: Southern California
Conventional least squares inversion

Inverse problem

Measurements $\rightarrow$ Model

"Measurement model"

$$ d = Am + n $$

"Inverse problem"

$$ \min_m \|d - Am\|^2 $$

(least-squares estimator)

Example model: Southern California
Background on LST methodology

- Most conventional travel time inversion methods regularize the inversion assuming exclusively smooth or discontinuous slownesses, potentially producing biased results.

- Earth contains both smooth and discontinuous variations in slowness at multiple spatial scales.

- Bianco and Gerstoft (2018) propose novel 2D travel-time tomography methods (LST) to get around the problem, featuring:
  - Sparsity constraints on slowness patches
  - Dictionary learning (unsupervised learning)
  - Damped least squares regularization on overall slowness map

Paper: "Travel time tomography with adaptive dictionaries"
Bianco and Gerstoft 2018, IEEE Transactions on Computational Imaging

Code online: https://github.com/mikebianco
Sparse models and dictionaries

• Sparse modeling assumes each signal model can be reconstructed from a few vectors from a large set of vectors, called a dictionary $D$

• Adds auxiliary sparse model to measurement model

\[ d = Am + n, \quad m \approx Dx \text{ and } |x| \ll Q \]

\[ \min_x ||ADx - d||_2 \quad \text{subject to } ||x||_0 \leq T \]

• Optimization changes from estimating $m$ to estimating sparse coefficients $x$
Dictionary learning and sparsity

• Obtains optimal sparse modeling dictionaries directly from data

• Assumes
  • (1) Redundancy in data: image patches are repetitions of a smaller set of elemental shapes
  • (2) Sparsity: each patch is represented with few atoms from dictionary

"Natural" images, patches shown in magenta
Each patch is signal, \( y = Dx \)

Learn dictionary \( D \) describing \( Y = [y_1, \ldots, y_I] \)

• Each patch is signal \( y_i \)
• Set of all patches \( Y = [y_1, \ldots, y_I] \)

Sparse model for patch \( y_i \) composed of few atoms from \( D \)

\[
y = \begin{bmatrix} \vdots \\
\end{bmatrix} = \begin{bmatrix} \vdots \\
\end{bmatrix} x_1 + \begin{bmatrix} \vdots \\
\end{bmatrix} x_2 + \ldots
\]
Checkerboard dictionary example

\[ y = R_i s = D x_i \]

10x10 pixel patches
\[ R_i \text{ selects image patch} \]

\[ x_1 + x_2 + \ldots \]

Olshausen 2009

Dictionaries

Slowness

Natural image
LST Basics: local and global models

Synthetic "checkerboard" slowness example

LST approach three ingredients: classified as **local** and **global** models

<table>
<thead>
<tr>
<th>“Local” model</th>
<th>1. Sparsity constraint on slowness patches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Dictionary learning (unsupervised machine learning)</td>
</tr>
<tr>
<td>“Global” model</td>
<td>3. Damped least squares regularization on overall slowness map</td>
</tr>
</tbody>
</table>

“Local” model: Models small-scale features as patches

“Global” model: Models larger-scale features with damped least squares
**LST Basics: local and global models**

Slowness map and sampling:
- Discrete slowness map $N = W_1 \times W_2$ pixels
- $I$ overlapping $\sqrt{n} \times \sqrt{n}$ pixel patches
- $M$ straight-ray paths

Tomography matrix (straight ray)

$$A \in \mathbb{R}^{M \times N}$$

Slowness dictionary

$$D \in \mathbb{R}^{n \times Q}, \quad Q \ll I$$

---

"Local" model

$$\hat{x}_i = \arg\min_{x_i} \| R_i s_s - Dx_i \|_2^2 \quad \text{subject to} \quad \| x_i \|_0 = T$$

---

"Global" model

$$t = As_g + \epsilon, \quad \hat{s}_g = \arg\min_{s_g} \| t - As_g \|_2^2 + \lambda_1 || s_g - s_s ||_2^2, $$

**Bayesian formulation**

Slowness map and measurements
- stations in red
- rays in blue
LST inversion of checkerboard, no noise (full)

LST inversion example (Bianco and Gerstoft 2018, IEEE TCI), Learned dictionary

LST Code available online: https://github.com/mikebianco
In March 2011, 5200 seismic stations were deployed in Long Beach, California (70km² area). Ambient seismic noise cross-correlations were obtained for all unique virtual source-receiver pairs (~14 million ray paths) using 3 weeks of data. We consider only the 1Hz vertical component data, corresponding to Rayleigh surface waves (from Lin et al. 2013). After quality control there were ~3 million ray paths.

"High-resolution seismic tomography of Long Beach, CA using machine learning" Bianco, Gerstoft, Olsen, Lin, Scientific Reports, 2019 (accepted).
1Hz LST phase speed map

- 300x200 pixel slowness map with 3 million rays (tomography matrix $A$ has dimensions $M=8$ million, $N=60000$)

- 10 iterations, used ~2 cpu-hours

"High-resolution seismic tomography of Long Beach, CA using machine learning" 
Bianco, Gerstoft, Olsen, Lin, Scientific Reports, 2019 (accepted).
Dictionary learned

- 200 atoms
- Atoms with sharper, oriented features correspond to sharper features in LST map
Groundwater resources threatened by seawater intrusion!
Stratigraphic interpretation

- Stratigraphy ~1 km east of LB array
- 1Hz Rayleigh surface wave sensitivity kernel (average) matches depth of Pleistocene "Silverado" aquifer

"Silverado" aquifer (Poland et al. 1956)
LST versus predicted Silverado 1Hz response

- Silverado estimated density $\sim 2,290 \text{ kg/m}^3$
- Surrounding layers 2,050-2,100 kg/m$^3$
- Density difference produces $\sim 2 \times V_s$ from empirical relations
Conclusions

- The proposed locally-sparse travel time tomography (LST) method with dictionary learning obtains potentially higher resolution than conventional methods.
- There is evidence we have imaged aquifer structure in Long Beach using LST.

"Travel time tomography with adaptive dictionaries"  
Bianco and Gerstoft 2018, IEEE Transactions on Computational Imaging

"High-resolution seismic tomography of Long Beach, CA using machine learning"  
Bianco, Gerstoft, Olsen, Lin, 2019 (Accepted).

LST Code available online: https://github.com/mikebianco

Acknowledgement:  
The authors would like to thank the ONR for funding this research.