A deep network for single-snapshot direction of arrival estimation

Emma Ozanich, Peter Gerstoft
Scripps Institution of Oceanography, UC San Diego

Haiqiang Niu
Institute of Acoustics, CAS

Introduction

In this paper, we reformulated conventional beamforming (CBF) as linear in the covariance matrix of the weights to motivate the use of the sample covariance matrix (SCM) as a feedforward neural network (FNN) input [1, 2]. A deep FNN was developed for estimating direction of arrival (DOA) of multiple incoherent sources.

Conventional beamforming

For measured acoustic data $\mathbf{p} \in \mathcal{C}^{N\times 1}$ on an N–element array, plane wave replicas $\mathbf{w} \in \mathcal{C}^{N\times 1}$ are often assumed,

$$w^n(\theta) = \frac{1}{\sqrt{N}} e^{i\frac{\omega}{c}(n-1)\ell\sin(\theta)}, \quad \mathbf{p} = \sum_{k=1}^{K} S_k \mathbf{w}(\theta_k) + \mathcal{CN}(0,\sigma^2) \tag{1}$$

$$\mathbf{W} = \mathbf{W}^{\mathrm{R}} + i\mathbf{W}^{\mathrm{I}} = \mathbf{w}\mathbf{w}^H, \;\; \mathbf{P} = \mathbf{P}^{\mathrm{R}} + i\mathbf{P}^{\mathrm{I}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{p}_l \mathbf{p}_l^H. \tag{2}$$

where $\mathbf{P}$ is estimated across L snapshots. Under this assumption, CBF can be written as linear in terms of the data sample covariance estimate $\mathbf{P}$ and weight covariance $\mathbf{W}$:

$$B(\theta) = \frac{1}{L}\sum_{l=1}^{L} |\mathbf{w}^H(\theta)\mathbf{p}_l|^2 = \frac{1}{L}\sum_{l=1}^{L} \mathrm{Tr}\{\mathbf{w}\mathbf{w}^H \mathbf{p}_l \mathbf{p}_l^H\} \approx \mathrm{Tr}\{\mathbf{W}^H\mathbf{P}\}, \tag{3}$$

$\mathbf{W}^{\mathrm{R}}$ and $\mathbf{W}^{\mathrm{I}}$ are the real and imaginary components of $\mathbf{W}$, with $\mathbf{P}^{\mathrm{R}}$ and $\mathbf{P}^{\mathrm{I}}$ for $\mathbf{P}$. $(\cdot)^H$ is the matrix complex conjugate transpose. $\mathbf{P}$ and $\mathbf{W}$ are Hermitian.

Feed–forward neural network

A feedforward neural network (FNN) is trained on T samples each with D features, $\mathbf{x}_t \in \mathcal{R}^{D\times 1}$, $t = 1, ..., T$. The training labels representing arrival angle are one-hot encoded. The hidden layer activations are rectified linear units (ReLu). The output function is a likelihood–like distribution over M classes,

$$y^j_{\mathrm{pred}} = f(\mathbf{w}^{j\top}\mathbf{x}_t), \quad j = 1, ..., \mathrm{M} \tag{4}$$

where $\mathbf{w}^j$ a weight vector. If $f(\cdot)$ is linear, FNN can be used to solve for $\mathbf{W}$ in (3), for example with Keras [3] software, by minimizing J,

$$\hat{\mathbf{w}}^m = \arg\min_{\mathbf{w}^m} \left\{ \sum_{t=1}^{T} (-J(\mathbf{y}_{t,\mathrm{true}}, \mathbf{y}_{\mathrm{pred}}(\mathbf{x}_t))) \right\} \tag{5}$$

The FNN model may be extended to a fully–connected model by adding hidden layers. Then $f()$ is the *softmax* function over M angles.

Simulations

Training, validation, and test sets were generated for two plane wave sources. The training data is noiseless ($\sigma^2 = 0$) and includes all combinations of θ_1, θ_2 for $\theta = [-90°, 90°), \Delta\theta = 1°$. A validation set and a test set were generated from 1000 Monte Carlo simulations with $\theta_1, \theta_2 \in \mathcal{U}\{-90°, 89°\}, \Delta\theta = 1°$. Gaussian random noise is added with signal–to–noise ratio (SNR)

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|\mathbf{p}_1\|_2^2}{\sigma^2}\right),$$
$$\sigma^2 = \|\mathbf{p}_1\|_2^2 \times 10^{-\frac{\mathrm{SNR}}{10}}. \tag{6}$$

Source incoherence

Plane wave sources with random phase, $S_k = |S|e^{i\phi_k}$, $\phi_k \in \mathcal{U}[-\pi, \pi]$, are incoherent if $\phi_j \neq \phi_k$, $j \neq k$. With few or single snapshots ($L = 1$), the SCM inputs to our FNN depend on the random source phases, ϕ_k. For example, for 2 sources,

$$P^{\mathrm{R}}_{n,m} = \frac{1}{N}\Big[\sum_{k=1}^{2} |S_k|^2 \cos\left(\frac{\omega}{c}(n-m)\ell\sin(\theta_k)\right) + 2S_1S_2\cos\left(\frac{\omega}{c}(n\sin(\theta_1) - m\sin(\theta_2))\ell + \Delta\phi\right)\Big] \tag{7}$$

where $n, m = 1, ..., N$ and $\Delta\phi = \phi_1 - \phi_2$. Table 1 compares FNN trained with coherent and incoherent sources. Each set contains 16110 samples.

Table 1: Accuracy for a test set with 1000 random simulations across SNR. FNN was trained on single coherent, single incoherent, and 5 incoherent training sets.

SNR	Coherent	Incoherent	Incoherent, 5 Sets
0 dB	0.30	0.34	0.21
10 dB	0.80	0.81	0.76
20 dB	0.95	0.93	0.94
50 dB	1	0.97	0.99
∞	1	0.86	0.98

At high SNR, increasing the number of incoherent training samples improves the incoherent DOA accuracy nearly to 1. These results indicate that the FNN learns to predict DOA from SCM features despite variation caused by the random source phase. Our training data is noiseless ($\sigma^2 = 0$), resulting in decreased performance at low SNR.

Hyperparameter selection

The validation set was used to compare models with different number of hidden layers (Table 2) in order to optimize the model for noiseless and noisy data.

Table 2: Validation accuracy vs number of FNN hidden layers, with 1024 hidden nodes.

Layers	0 dB	10 dB	20 dB	100 dB	Training
1	0.22	0.48	0.68	0.99	1.0
2	0.17	0.60	0.87	0.99	1.0
3	0.22	0.0.7	0.92	0.99	1.0
4	0.22	0.71	0.93	1.0	1.0
5	0.24	0.70	0.93	1.0	1.0
6	0.17	0.68	0.92	1.0	1.0
7	0.14	0.56	0.89	1.0	1.0
8	0.14	0.57	0.88	1.0	1.0

At high SNR, increasing the model depth (hidden layers) improves DOA accuracy indefinitely. At lower SNR (<20 dB), DOA accuracy decreases for FNN models deeper than 5 layers. Based on these results, we apply the FNN with 5 hidden layers to unseen test data.

Accuracy

The accuracy between the estimated peaks, $\hat{\theta}_t$ and the true angles $\theta_{t,j}$ is

$$\mathrm{Accuracy} = \frac{1}{K\cdot T}\sum_{j=1}^{K}\sum_{t=1}^{T} \mathbf{1}\Big[\Big|\theta_{t,j} - \hat{\theta}^j_t\Big| \leq 1°\Big] \tag{8}$$

where $\mathbf{1}[x]$ is the indicator function.

Notation

n: array index, S_k: source amplitude, K: number of sources
ω: source frequency, c: phase speed, ℓ: array spacing
$\omega = (2\pi \times 200)$ rad/s, $c = 1500$ m/s, $\ell = 2.5$ m, $N = 20$ elements.

Two source direction of arrival with deep network

A 5–layer deep FNN trained with 5 incoherent training sets was used to track two noiseless, incoherent source tracks (Fig. 1a). The FNN had 512 hidden nodes per layer and was trained for 1000 epochs.

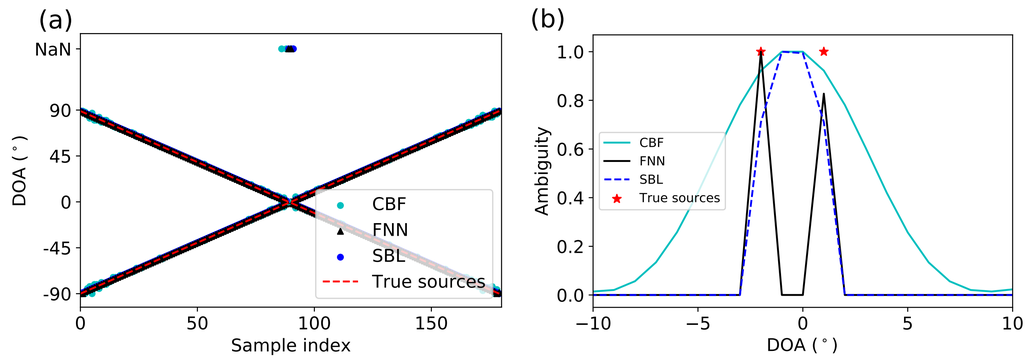

Figure 1: *(a)* Test set DOA estimation of two noiseless incoherent sources from two–peak detection along two simultaneous source tracks. Missing detections (NaNs) are shown at the top of the grid. *(b)* Ambiguities of the two–peak detection at $\theta_1 = 1°, \theta_2 = -2°$.

The FNN trained on 5 incoherent training sets had a high accuracy for tracking 2 incoherent sources with high SNR. CBF and sparse Bayesian learning (SBL) [4] were compared to FNN for the single–snapshot scenario. At the source track crossing, FNN distinguishes both peaks, while CBF and SBL fail for single snapshot (Fig. 1b). A NaN indicates that only one of two peaks were predicted.

Multisource direction of arrival with deep network

In realistic applications the number of sources will not be known *a priori*. We thus retrained the FNN model to predict a random number of incoherent sources by randomly selecting the training data.

We generated 10^6 noiseless, incoherent training samples. Each sample contained K plane wave sources at angles θ_k, with $K \in \mathcal{U}[1, 10]$ and $\theta_k \in [-90°, 90°), \Delta\theta = 1°$.

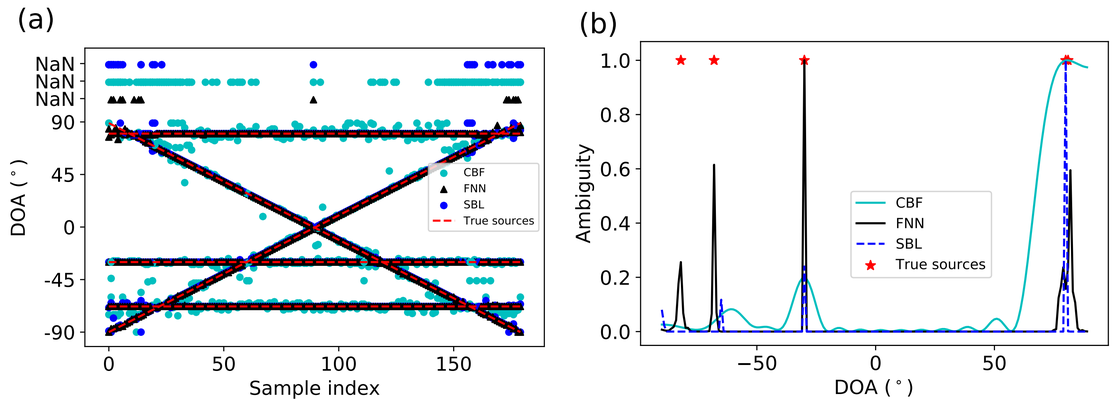

Figure 2: *(a)* Test set DOA estimation of five noiseless, incoherent sources for FNN trained on a random number of sources. *(b)* Ambiguities of the five–peak detection at $\theta_1 = -82°, \theta_2 = 81°, \theta_3 = -68°, \theta_4 = -30°$, and $\theta_5 = 80°$.

The randomly trained FNN was tested on tracks of 5 noiseless, incoherent sources and compared to CBF and SBL (Fig. 2), with source angles $\theta_1 = [-90°, ..., 89°]$, $\theta_2 = [89°, ..., -90°]$, $\theta_3 = [-68°, ..., -68°]$, $\theta_4 = [-30°, ..., -30°]$, $\theta_5 = [80°, ..., 80°]$. FNN achieved **89%** accuracy while SBL achieved **83%**.

Experimental two source direction of arrival

As an experimental test, we estimated the azimuth of the Swellex–96 S95 deep source tow and a loud interferer (Fig. 3) to the North horizontal line array [5, 6]. The North array was arranged in a slight arc, allowing for left–right discrimination by high resolution methods. The true array positions were used to compute the weights for FNN training data, SBL, and CBF. The SCM was constructed at 79 Hz, with $\Delta f = 0.8$ Hz.

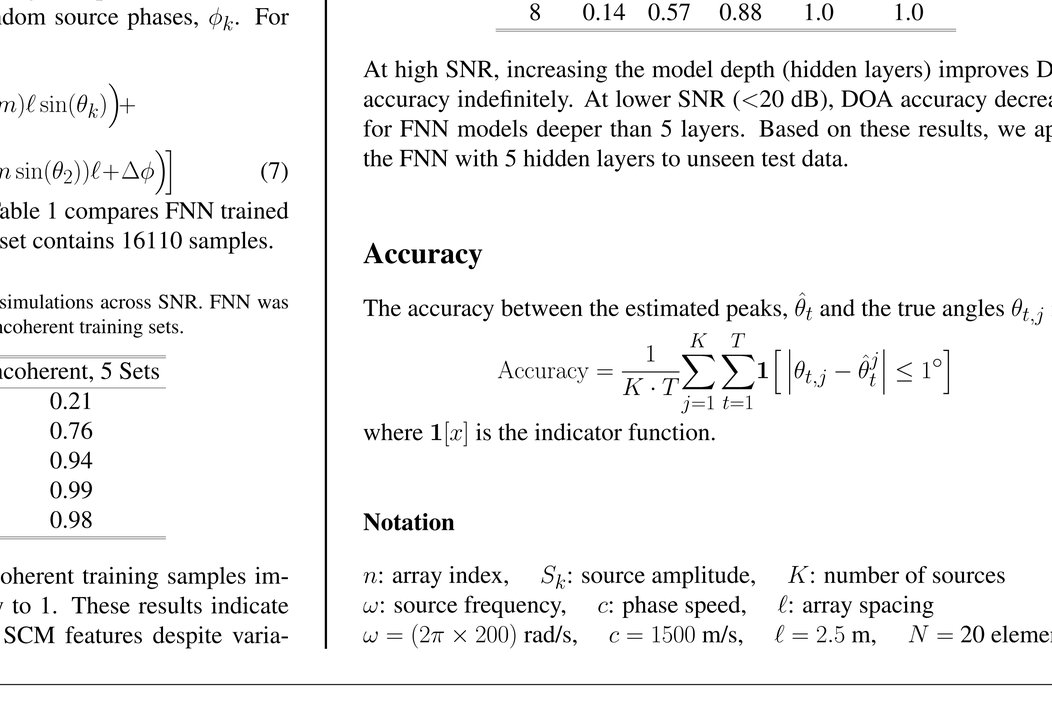

Figure 3: $L = 10$ snapshot *(a)* FNN, *(b)* SBL, and *(c)* CBF ambiguities for the Swellex–96 S95 source tow event (circles) and a loud interferer (stars). Array endfire (solid) and broadside (dashed) directions are shown. SBL and FNN were convolved with a 3x3 unit filter for improved visualization.

Figure 3 shows the ambiguity surfaces for FNN, SBL, and CBF, normalized to their maximum at each time. In total, there were 3120 test samples, with the SCM averaged over $L = 10$ snapshots. FNN achieved resolution similar to SBL, demonstrating its potential in experimental applications.

References

1. Niu, H. et al. *J. Acoust. Soc. Am.* **142,** 1176–1188 (2017).
2. Niu, H. et al. *J. Acoust. Soc. Am.* **142,** EL455–EL460 (2017).
3. Chollet, F. et al. *Keras* https://keras.io. 2015.
4. Nannuru, S. et al. *J. Acoust. Soc. Am.* **144,** 2719–2729 (2018).
5. Hursky, P. et al. *J. Acoust. Soc. Am.* **109,** 1355–1366 (2001).
6. Gemba, K. et al. *J. Acoust. Soc. Am.* **141,** 92–103 (2017).