Learnable kernel-based FRI reconstruction
Bachelor Thesis Project 2

Omkar Nitsure
Guide: Prof. Satish Mulleti

Electrical Engineering, IIT Bombay

May 2, 2025

Github Link:
https://github.com/omkar-nitsure/Learnable_kernel_FRI/tree/main

Introduction

- Finite Rate of Innovation (FRI) sampling enables efficient reconstruction of sparse signals
- FRI models allow sub-Nyquist sampling
- FRI signals appear in applications such as:
  - Radar, LIDAR, OCT (Optical Coherence Tomography), EEG, ECG, Medical imaging, Source localization
- FRI signals have linear combinations of delayed, scaled versions of a known pulse
- Goal: Estimate amplitudes and delays from a few noisy measurements

FRI Signal Model

- Consider signals of the form:

$$f(t) = \sum_{\ell=1}^{L} a_\ell \, h(t - \tau_\ell)$$

  where:
  - $h(t)$: known FRI pulse
  - a_ℓ: amplitudes in $[a_{\min}, a_{\max}]$
  - τ_ℓ: delays in $[\tau_{\min}, \tau_{\max}]$, sorted
- Known model order: L
- $h(t)$ has compact support: $[T_{h,\min}, T_{h,\max}]$

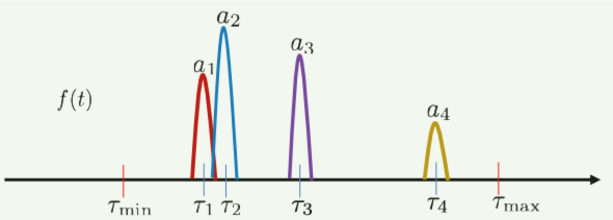


Figure 1: FRI signal setup

Need for Filtering

- Direct sampling of $f(t)$ requires high rate when $h(t)$ is wideband
- Solution: Use a sampling kernel $g(t)$ with larger support
- Convolution with $g(t)$ broadens $f(t)$ and reduces required sampling rate
- Enables sub-Nyquist sampling of $f(t)$

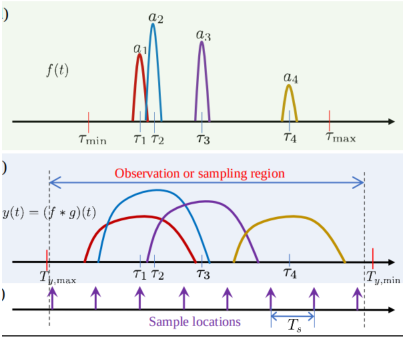


Figure 2: Filtering enables lower-rate sampling

Sum-of-Exponentials Model and Recovery

- Filtered samples $\rightarrow$ linear measurements:

$$z(m) = \sum_{\ell=1}^{L} b_\ell \, u_\ell^m, \quad m = 0, \ldots, M-1$$

- Requires $M \geq 2L$ for exact recovery
- $u_\ell = e^{j\omega_0 \tau_\ell}$ (example)
- Recovery methods:
  - Annihilating filter, ESPRIT, Matrix Pencil, etc.
  - Sensitive to noise

Noise Robustness and Resolution

- Classical methods degrade in noise, especially at small $\Delta\tau$:

$$\Delta\tau = \min_{\ell} |\tau_{\ell+1} - \tau_{\ell}|$$

- Denoising techniques (e.g., Cadzow) used before parameter recovery
- Increasing the number of samples improves noise-robustness
- Sequential methods reduce complexity but have low resolution (depends on the filter decay)

Learning-Based Approaches

- Joint filter design + recovery via learning
- Learning-based reconstruction of FRI signals (Leung et al.):
  - Uses autoencoder for off-grid recovery
  - First training step - Trains the encoder to predict locations
  - Second training step - Trains the encoder (with frozen decoder for known kernel) to jointly optimize location prediction and signal reconstruction
  - Improves resolution and noise robustness
- Still relies on exponential-generating kernels

Proposed Framework

- FRI signal $f(t)$ is filtered with a learnable kernel $g_{\boldsymbol{\theta}}(t)$
- Samples $\mathrm{y}_{\boldsymbol{\theta}} = (f * g_{\boldsymbol{\theta}})(t)$
- A deep encoder E_ϕ estimates the time delays:

$$\hat{\boldsymbol{\tau}} = E_\phi(\mathrm{y}_{\boldsymbol{\theta}})$$

- Amplitudes are then estimated separately using $\hat{\boldsymbol{\tau}}$ and $\mathrm{y}_{\boldsymbol{\theta}}$ (similar to least-squares learning)

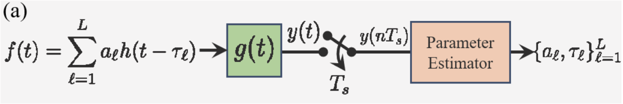


Figure 3: Flowchart

Training the Encoder

- Fix kernel parameters $\boldsymbol{\theta}$
- Train encoder using database $\mathcal{D}_{\text{train}} = \{\mathrm{y}_{\boldsymbol{\theta},i}, \boldsymbol{\tau}_i\}_{i=1}^{I}$
- Optimization objective:

$$\min_{\phi} \sum_{i=1}^{I} \|\boldsymbol{\tau}_i - E_{\phi}(\mathrm{y}_{\boldsymbol{\theta},i})\|_p^p$$

- Choice of $p = 1$ or 2 for loss
- We empirically found out that the L1-loss works better than the L2-loss probably because it induces sparsity

Joint Learning of Kernel and Encoder

- Learn both $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ jointly:

$$\min_{\boldsymbol{\theta},\boldsymbol{\phi}} \sum_{i=1}^{I} \|\boldsymbol{\tau}_i - E_{\boldsymbol{\phi}}(\mathrm{y}_{\boldsymbol{\theta},i})\|_p^p$$

- Backpropagation used to update both:

$$\boldsymbol{\phi}^{(k+1)} = \boldsymbol{\phi}^{(k)} - \eta_\phi \sum_i \nabla_\phi \mathcal{L}_i$$

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \eta_\theta \sum_i \nabla_\theta \mathcal{L}_i$$

Amplitude Estimation

- Amplitudes estimated after delay prediction:

$$\hat{\mathrm{a}} = \arg\min_{\mathrm{a}} \frac{1}{N} \sum_{n=1}^{N} \left| \hat{y}(nT_s; \mathrm{a}, \hat{\boldsymbol{\tau}}, g_{\boldsymbol{\theta}}) - \mathrm{y}_{\boldsymbol{\theta}}[n] \right|^2$$

- Solved via gradient descent (We use the Adam optimizer):

$$\mathrm{a}^{(k+1)} = \mathrm{a}^{(k)} - \eta \nabla_{\mathrm{a}} \mathcal{L}_{\mathrm{MSE}}(\mathrm{a}^{(k)})$$

- Separating delay and amplitude recovery improves accuracy and stability (while allowing the parameter size of the model to be reduced significantly)

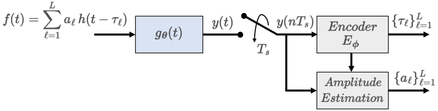


Figure 4: Block diagram

Learning Reconstruction with Arbitrary Kernels

Goal:

- Learn signal reconstruction methods from samples generated using arbitrary kernels
- We demonstrate that Sum of Exponentials (SoE) generation is **not necessary** for accurate reconstruction

Approach:

- We used truncated Gaussian / Gaussian pair as sampling kernels (these are even used as initialization of the learnable kernel)
- Avoided SoE-based reconstruction to circumvent instability issues in noisy settings

Comparison:

- Benchmarked against FRIED-Net

FRI Data Generation

Signal Model:

- $f(t) = \sum_{\ell=1}^{L} a_\ell \delta(t - \tau_\ell)$ (stream of Diracs)
- $a_\ell \sim \mathcal{U}[0.5, 10], \quad \tau_\ell \sim \mathcal{U}[-0.476, 0.5231]$

Sampling Setup:

- Convolve with compactly supported kernel $g_{\boldsymbol{\theta}}(t)$
- Sample uniformly over $[-0.9, 0.9]$ with $T_s = 0.086$ ($N = 21$ samples)
- Kernel support: $[T_{g,\min}, T_{g,\max}] = [-0.3, 0.3]$

Key Differences from FRIED-Net:

- FRIED-Net uses a kernel spanning the full observation window
- Our truncated kernel avoids boundary interference and enhances robustness

Additional Experiments:

- **Low-rate Sampling:** $T_s = 0.16$, $N = 11$ samples
- **Higher Pulse Counts:** $L = 5$ and $L = 10$

Kernel Types Evaluated

Truncated Gaussian Kernel

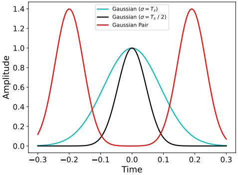
$$g(t) = e^{-\frac{t^2}{2\sigma^2}}, \quad t \in [-0.3, 0.3]$$

- Empirically: $\sigma \geq T_s/2$ is effective (best results for $\sigma = T_s/2$)

Gaussian Pair Kernel

$$g(t) = Ae^{-\frac{(t+t_1)^2}{2\sigma^2}} + Be^{-\frac{(t+t_2)^2}{2\sigma^2}}$$

- Improves resolution for closely spaced pulses



Results: Comparison with FRIED-Net

- Objective: Evaluate resolution performance under varying noise levels
- Baseline: FRIED-Net [5]
  - Uses eMOMS kernel with infinite support over $[-0.476, 0.5231]$
- Ours: Compact kernels (Gaussian, narrow Gaussian, Gaussian pair) with support $[-0.3, 0.3]$
- Same number of samples $N = 21$, different sampling intervals:
  - FRIED-Net: $T_s = 0.047$
  - Ours: $T_s = 0.086$
- Wider observation window: $[-0.9, 0.9]$

Key Results: Gaussian Kernels

Gaussian Kernel ($\sigma = T_s$)

- Balances smoothness and resolution
- 5–7 dB improvement at $\Delta\tau = 0.05$ vs. FRIED-Net (high SNR)

Narrow Gaussian ($\sigma = T_s/2$)

- Improves resolution for closely spaced pulses
- Average 6 dB gain at $\Delta\tau = 0.05$ (25–40 dB SNR)

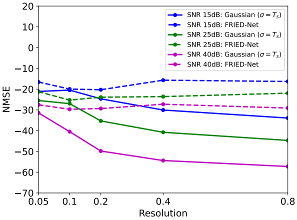


Figure 5: Gaussian ($\sigma = T_s$)

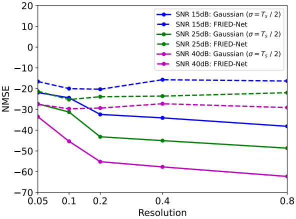
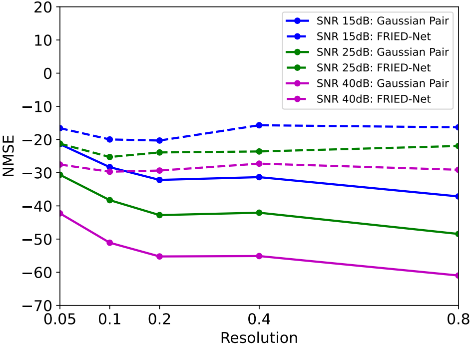


Figure 6: Gaussian ($\sigma = \frac{T_s}{2}$)

Key Results: Gaussian Pair Kernel

- Best overall performance across settings
- Dual peaks enhance local sensitivity
- Robust to noise; 7–8 dB gain at $\Delta\tau = 0.05$
- Resolves closely spaced pulses more effectively than FRIED-Net



Kernel Parameterization

- The sampling kernel $g_{\boldsymbol{\theta}}(t)$ is parameterized using first-order B-spline basis functions:

$$g_{\boldsymbol{\theta}}(t) = \sum_{k=-K}^{K} c_k \, \beta_1 \left(\frac{t - kT}{T} \right)$$

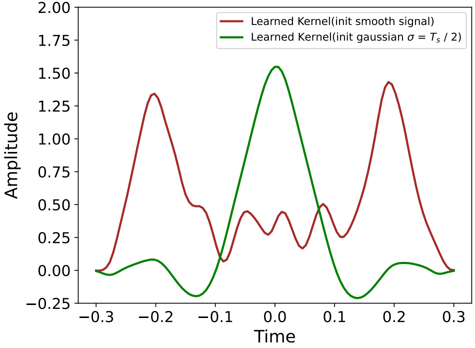
- The first-order B-spline $\beta_1(t)$ is defined as:

$$\beta_1(t) = \begin{cases} 1 - |t| & \text{if } |t| \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

- This construction results in a piecewise linear kernel with compact support
- The kernel is learnable, and the model can adapt its shape during training

Training Method for Jointly Learned Kernel

- Two training configurations:
  1. **Learned Kernel Initialized with Smooth Function:**
     - Kernel coefficients are initialized with a smooth function
     - Optimized with ℓ_2-norm loss
  2. **Learned Kernel Initialized with Gaussian Function:**
     - Kernel initialized to approximate a Gaussian function
     - Optimized with ℓ_1-norm loss for robustness to outliers



Results

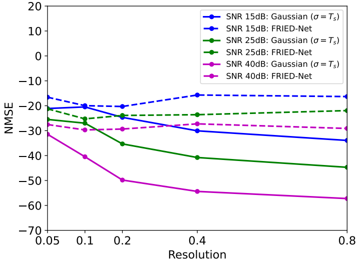


Figure 7: Smooth Initialization

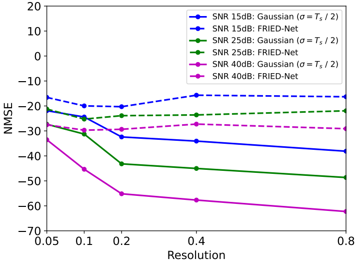


Figure 8: Gaussian $(\sigma = \frac{T_s}{2})$

Resolution Analysis

- Test on 1000 examples focusing on pulse separation ($\Delta\tau = 0.05$ to 0.8)
- FRIED-Net uses an eMOMS kernel with full observation window $[-0.476, 0.5231]$
- Our method uses compactly supported learned kernels with support $[-0.3, 0.3]$
- Performance improvement:
  - 5–7 dB NMSE improvement for smooth-initialized kernel
  - Up to 9 dB improvement with Gaussian-initialized kernel at higher SNRs
- Conclusion: The joint learning framework enhances resolution, even with compact kernels

Generalization Accuracy

- Tested on 1000 randomly generated FRI signals
- Performance compared in terms of NMSE for time delay (t_k) and amplitude (a_k) estimation
- Learned Gaussian kernel outperforms fixed kernels in both time delay and amplitude estimation
- At low SNRs (15 dB), the learned Gaussian kernel shows:
  - 3–4 dB gain in t_k estimation
  - 2–3 dB gain in a_k estimation
- Conclusion: Learned kernel provides superior generalization and accuracy across various signal conditions

Reduced Sampling Rate

- Evaluated learned kernel initialized with a Gaussian function under reduced sampling rate ($N = 11$)
- Despite halving the sampling rate, the model performs well, with only a 6–8 dB degradation in NMSE compared to $N = 21$
- Performance surpasses several fixed-kernel approaches at $N = 21$
- Conclusion: The framework is robust under sparse sampling, making it suitable for resource-constrained applications

Hardware Implementation

- We investigate the practical implementation of the learned kernel using a 2-pole system
- The pole and gain values are optimized for accurate location reconstruction
- Emphasis on maximizing the pole magnitude governing exponential decay:
  - Results in less spread-out kernels, improving resolution accuracy
  - Simplifies hardware realization
- In the lab, a synthetic pipeline is followed:
  - Discrete samples fed to DAC to generate a continuous analog signal
  - Signal passed through filter, then captured by ADC
  - Captured signal is fed to the model as input
- Conclusion: The learned kernel paradigm enables the synthesis of optimal kernels, achieving excellent reconstruction accuracy

Filter Specifications

- All previous simulations use Dirac impulses for FRI signal generation, but practical signal generators can only generate pulses with finite bandwidth. Thus, we chose the FRI pulse bandwidth to be 1 kHz
- We design a 2-pole Opamp-based filter and achieve comparable performance to a learnable filter with no specific design requirements (around 5-6 dB performance drop)

The filter can be represented as follows -

$$H(s) = \frac{a_1 a_2}{(s + \beta_1)(s + \beta_2)} \tag{1}$$

$$h(t) = a\left(e^{-\beta_1 t} - e^{-\beta_2 t}\right) = \frac{a_1 a_2}{\beta_2 - \beta_1}\left(e^{-\beta_1 t} - e^{-\beta_2 t}\right) \tag{2}$$

$\frac{a_1 a_2}{\beta_2 - \beta_1} = 3.68$, $\beta_1 = 43.53$, $\beta_2 = 61.07$

Filter responses

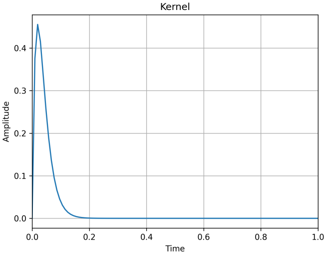


Figure 9: Time domain response

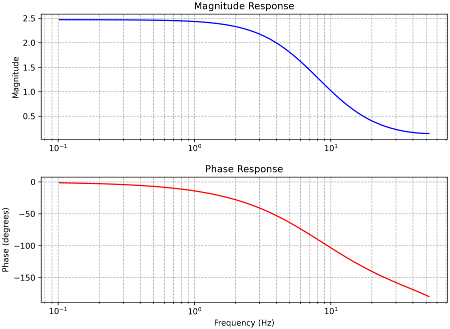
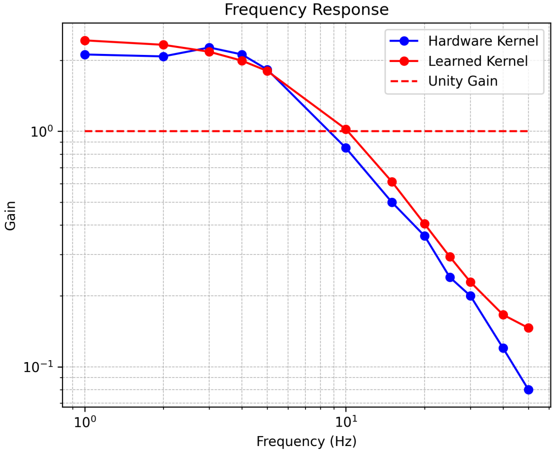


Figure 10: Frequency response

Comparison of hardware realization with Simulated kernel



References I

[1] J. A. Cadzow. "Signal enhancement - A composite property mapping algorithm". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36.1 (Jan. 1988), pp. 49–62. ISSN: 0096-3518. DOI: 10.1109/29.1488.

[2] G. R. DeProny. "Essai experimental et analytique: Sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool, à différentes températures". In: *J. de l'Ecole polytechnique* 1.2 (1795), pp. 24–76.

[3] Y. Hua and T. K. Sarkar. "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise". In: *IEEE Trans. Acoust., Speech and Signal Process.* 38.5 (May 1990), pp. 814–824. ISSN: 0096-3518. DOI: 10.1109/29.56027.

References II

[4] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[5] Vincent CH Leung et al. "Learning-based reconstruction of FRI signals". In: *IEEE Trans. Signal Process.* 71 (2023), pp. 2564–2578.

[6] Satish Mulleti, Haiyang Zhang, and Yonina C Eldar. "Learning to sample: Data-driven sampling and reconstruction of FRI signals". In: *IEEE Access* 11 (2023), pp. 71048–71062.

[7] A. Paulraj, R. Roy, and T. Kailath. "A subspace rotation approach to signal parameter estimation". In: *Proc. IEEE* 74.7 (1986), pp. 1044–1046. ISSN: 0018-9219. DOI: 10.1109/PROC.1986.13583.

Questions/Suggestions

Thank you!