Neural Stimulation Reconstruction from EEG using Fractional-Order Networks towards Predictive Model Validation in Clinical Applications

Alessandro Varalda and Sergio Pequito

***Abstract*—Effective therapeutic neurostimulation requires predictive models that can reliably map neural activity responses to specific stimuli. While recent advances in closed-loop neurostimulation devices show promise for treating neurological disorders, most modeling approaches neglect the crucial relationship between stimulation input and neural response.**

A fundamental test of model predictive capability remains unaddressed: given baseline electroencephalographic (EEG) data and subsequent neural responses to stimulation, can we accurately reconstruct information about the injected stimuli? Furthermore, can such reconstruction remain valid across different stimulation parameters and electrode configurations, reflecting the diverse electrode placements required in clinical practice?

In this paper, we demonstrate that input reconstruction is achievable using discrete-time linear fractional-order dynamical networks (DTLFON), which capture the rich temporal dependencies characteristic of neural systems. We present a novel minimization-minimization algorithm that generalizes expectation-maximization principles to learn both system parameters and unknown inputs.

Our approach differs from previous methods by explicitly incorporating a learning phase to establish baseline dynamics before stimulation, allowing for more reliable parameter estimation while accounting for the quasi-stationary nature of stimulation electrode configurations.

I. Problem Formulation and Methodology

In neural recordings, we observe voltage measurements $x[k] \in \mathbb{R}^n$ at $n \in \mathbb{N}$ different spatial locations through EEG readings at discrete time steps $k \in \mathbb{N}$, reflecting both intrinsic brain dynamics and responses to external stimuli $u[k] \in \mathbb{R}^p$ with $p \in \mathbb{N}$.

We model the input-output relationship using a discrete-time linear fractional-order dynamical network (DTLFON):

$$\Delta^\alpha x[k+1] = Ax[k] + Bu[k] + w[k], \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ captures spatial coupling between EEG electrode readings, $B \in \mathbb{R}^{n\times p}$ describes how inputs influence these readings, $w[k] \in \mathbb{R}^n$ represents process noise, and Δ^α represents a fractional difference operator of order $\alpha \in \mathbb{R}^n$.

This operator captures different time-scales of neural activity through coefficients $\psi(\alpha_i, j)$, where $\psi(\alpha_i, j) = \Gamma(j - \alpha_i)/[\Gamma(-\alpha_i)\Gamma(j+1)]$ and $\Gamma(\cdot)$ is the Gamma function. These coefficients control how strongly past states influence current dynamics, with higher α_i values corresponding to stronger memory effects, while lower values indicate more rapid decay of historical influence.

Unlike conventional data-driven approaches that focus solely on predicting neural activity patterns, our research addresses the inverse problem: can we accurately reconstruct both the unknown input sequence $\{u[k]\}_{k=0}^{T-1}$ and the underlying network structure that maps these inputs to the observed neural responses?

This inverse problem presents several key challenges: (i) the nonlinear relationship between measurements and unknown parameters, (ii) measurement noise and system uncertainties, (iii) long-memory effects captured by the fractional-order dynamics, and (iv) coupling between spatial and temporal dynamics in the network. Additionally, the problem is inherently ill-posed, as multiple combinations of inputs and network parameters could potentially explain the observed measurements.

To solve this inverse problem, we develop a novel minimization-minimization (min-min) algorithm that alternates between estimating system parameters and reconstructing unknown inputs. The algorithm consists of two complementary optimization stages:

- **Stage 1: Gradient Optimization for Temporal-Spatial Parameters** - During this phase, we assume no external inputs ($u[k] = 0$) and estimate system parameters α and A by implementing gradient descent optimization that minimizes prediction error. This stage begins by initializing fractional orders $\alpha_i^{(0)} = 0.5$ for all state variables with $i \in \{1, 2, \ldots, n\}$. For each iteration $l \in \mathbb{N}$, we compute fractional derivatives $z_i[k] = \sum_{j=0}^{J-1} \psi(\alpha_i^{(l)}, j)x[k+1-j]$ with a truncation term $J \in \mathbb{N}$ that limits the infinite sum. We then solve for the optimal coupling matrix $A^{(l)}$ using least squares optimization: $A^{(l)} = \arg\min_A \|Z - AX\|^2$, where $Z \in \mathbb{R}^{n\times(k+1)}$ and $X \in \mathbb{R}^{n\times(k+1)}$ are matrices of stacked derivatives and states, respectively. The fractional orders are updated through quasi-Newton optimization: $\alpha^{(l+1)} = \alpha^{(l)} - \eta\nabla_\alpha e^{(l)}$, with $\eta \in \mathbb{R}^+$ being the learning rate and $e^{(l)} \in \mathbb{R}$ the estimation error. This process continues until convergence.
- **Stage 2: Input Reconstruction** - After establishing baseline dynamics, we reconstruct both the input coupling matrix B and unknown input sequence $u[k]$ through alternating minimization steps. We compute residuals $r[k] = z[k] - Ax[k] \in \mathbb{R}^n$, which represent the part of the dynamics unexplained by the network's intrinsic behavior.

  Given the current estimate $B^{(l)} \in \mathbb{R}^{n\times p}$, we reconstruct the input sequence $\hat{u} = \arg\min_u \|R - B^{(l)}u\|^2$ (first

minimization), where $R \in \mathbb{R}^{n\times(k+1)}$ is the matrix of stacked residuals. Then, we update the coupling matrix $B^{(l+1)} = [b_1^{(l+1)}, ..., b_n^{(l+1)}]^T$, where each element $b_i^{(l+1)} = \arg\min_b \|R_i - b\hat{u}\|^2$ (second minimization) with $R_i \in \mathbb{R}^{1\times(k+1)}$ denoting the i-th row of the residual matrix R.

The algorithm alternates between these steps until convergence, addressing the inherently ill-posed nature of this inverse problem.

II. Experimental Validation and Results

We validated our approach through both numerical simulations with synthetic data and analyses of real clinical recordings. This two-pronged strategy allowed us to verify algorithm performance under controlled conditions, assess robustness to realistic noise and variability, test generalizability across different stimulation patterns, and compare performance with traditional linear time-invariant (LTI) approaches.

We evaluated reconstruction quality using multiple complementary performance indicators, each capturing different aspects of reconstruction fidelity. Dynamic Time Warping (DTW) distance $\in \mathbb{R}^+$ was used to evaluate temporal alignment between true and reconstructed signals, with values approaching zero indicating optimal alignment. Relationship strength was assessed through Pearson correlation coefficient (pCorr) $\in [-1, 1]$ for linear relationships and Spearman correlation coefficient (sCorr) $\in [-1, 1]$ for monotonic relationships, with values closer to 1 indicating optimal reconstruction.

Signal-to-Noise Ratio (SNR) $\in \mathbb{R}$ measured reconstruction quality in terms of signal power in decibels (dB), with higher values (above 20dB) indicating excellent quality. Normalized Root Mean Square Error (NRMSE) $\in \mathbb{R}^+$ quantified reconstruction accuracy independent of signal magnitude, with values below 0.1 indicating good reconstruction.

Finally, the coefficient of determination (R^2) $\in (-\infty, 1]$ assessed how well the reconstructed signal captured the dynamic range of the true input, with values above 0.8 suggesting excellent reconstruction quality.

In simulations using a synthetic four-state system, our DTLFON-based method consistently outperformed the LTI approach across multiple input types (sinusoidal, square wave, and biphasic pulse) and noise levels. Even under high noise conditions ($\sigma = 0.64$ with $\sigma \in \mathbb{R}^+$), the DTLFON approach maintained strong performance (Pearson correlation = 0.98, SNR = 7.9 dB, $R^2 = 0.84$) while the LTI method degraded significantly (Pearson correlation = 0.06, SNR = -6.3 dB, $R^2 = -3.30$).

For clinical validation, we used data from Mikulan et al. [2] featuring simultaneous recordings of intracerebral electrical stimulation and high-density EEG from epilepsy patients with approximately 110 valid trials recorded using a 256-channel HD-EEG system. Stimulation consisted of controlled biphasic pulses delivered through intracerebral electrodes at 14 different anatomical locations with intensities ranging from 0.1 to 5 mA.

Our approach demonstrated remarkable consistency across different stimulation locations (K13-14, N2-3, S1-2, S3-4) and intensities (1mA to 5mA). Notably, the method showed particular robustness in handling different array dimensionalities ($n \in \{1, 2, 4, 8\}$) while maintaining reconstruction quality, a critical feature for practical clinical deployment.

The superior performance of our DTLFON approach compared to traditional LTI methods was particularly evident in high-noise conditions typical of clinical settings. This advantage stems from the DTLFON method's ability to capture the complex temporal dependencies inherent in neural signals, providing better noise rejection capabilities and more accurate reproduction of the signal's dynamic range. The most significant performance differences were observed in the SNR and R^2 metrics, which specifically measure signal quality and variance explained, respectively.

For parameter estimation accuracy assessment, we used NRMSE between true and estimated values for system parameters (A, B, α). Our method achieved near-perfect parameter reconstruction under low noise conditions, with only modest degradation as noise increased. This exceptional parameter estimation capability underpins the superior input reconstruction performance demonstrated in our evaluation metrics.

III. Conclusion

Our research addresses a fundamental challenge in therapeutic neurostimulation: reliably predicting and validating neural responses to stimulation inputs. We demonstrated that discrete-time linear fractional-order networks can effectively reconstruct stimulation patterns from EEG measurements, providing a crucial step toward predictive model validation in clinical applications [1]. These results establish a promising foundation for advancing personalized therapeutic interventions, where accurate reconstruction of neural responses to stimulation could enable more precise and patient-specific treatment strategies.

The fractional-order approach [3], [4] provides significant advantages over traditional methods, especially in handling the complex temporal dependencies and noise characteristics typical of neural systems. As closed-loop neurostimulation devices continue to evolve, our method offers a valuable tool for both testing predictive model efficacy and optimizing stimulation parameters for individual patients.

References

[1] D. R. Nair, K. D. Laxer, P. B. Weber, A. M. Murro, Y. D. Park, G. L. Barkley, et al., "Nine-year prospective efficacy and safety of brain-responsive neurostimulation for focal epilepsy," *Neurology*, vol. 95, no. 9, pp. e1244-e1256, 2020.

[2] E. Mikulan, S. Russo, S. Parmigiani, et al., "Simultaneous human intracerebral stimulation and HD-EEG, ground-truth for source localization methods," *Scientific Data*, vol. 7, p. 127, 2020.

[3] D. Baleanu, J. A. T. Machado, and A. C. J. Luo, "Fractional Dynamics and Control," Springer Publishing Company, Incorporated, 2011.

[4] E. Reed, S. Chatterjee, G. Ramos, P. Bogdan, and S. Pequito, "Fractional cyber-neural systems–A brief survey," *Annual Reviews in Control*, vol. 54, pp. 386-408, 2022.