ECE 516: Adaptive Digital Filters
Lecture 19 (Selected Topics: Adaptive Filtering by Machine Learning)
Over the past several years, data-driven methods, and specifically deep learning techniques, have attracted unprecedented attention from research communities, across the board. The advent of low-cost specialized powerful computing resources (e.g., GPUs, and more recently TPU) and the continually increasing amount of massive data generated by the human population and machines, in conjunction with the new optimization and learning methods, have paved the way for deep neural networks (DNNs) and other machine learning-based models to prove their effectiveness in many engineering areas.
Parametrized mathematical models play a central role in understanding and design of complex information systems that are becoming common in our age. However, they often cannot take into account the intricate interactions innate to such systems. On the contrary, purely data-driven approaches do not need explicit mathematical models for data generation and have a wider applicability at the cost of interpretability. Although the data-driven approaches can handle large and complex datasets, they are ignorant to the underlying mathematical model of the systems generating them. Thus, it is vital to develop hybrid data-driven and model-based frameworks to enhance the accuracy, computational complexity, and efficiency of the data acquisition model in complex, large-scale scenarios.
Is there an intuitive way to combine the classical model-based statistical signal processing methods with data-driven models?

Can we do **model-based deep learning**?
Is there an intuitive way to combine the classical model-based statistical signal processing methods with data-driven models?

Can we do **model-based deep learning**?  Yes.
The opportune moment for model-based deep learning. The recent advent of the deep unfolding framework has paved the way for a game-changing fusion of models, and well-established signal processing approaches, with data-driven architectures. In this way, we not only exploit the vast amounts of available data but also integrate the prior knowledge of the system model in the processing stage. Deep unfolding networks (DUNs) rely on the establishment of an optimization or inference iterative algorithm, whose iterations are then unfolded into the layers of a deep network, where each layer is designed to resemble one iteration of the optimization/inference algorithm. The proposed hybrid method benefits from the low computational cost (in execution stage) of deep neural networks, and at the same time, from the flexibility, versatility, and reliability of model-based methods. Most importantly, the emerging networks appear to be an excellent tool in scalable machine learning applications due to the smaller degrees of freedom required for training and execution (afforded by the integration of the problem-level reasoning, or the model, see Fig. 1). Therefore, there is an untapped potential for deep unfolding to breed a new generation of deep signal processing architectures that are interpretable, and extremely appropriate for large-scale scenarios.
Deep Unfolding-Based Scalable Machine Learning

**General DNNs:**
Massive networks, difficult to train in real-time or large-scale settings.

**Deep Unfolding (DUNs):**
Incorporating problem level reasoning (models) in the deep network architecture, leading to sparser networks amenable to scalable machine learning.

*Figure 1: General DNNs vs DUNs. DUNs appear to be an excellent tool in scalable machine learning applications due to the smaller degrees of freedom required for training and execution.*
Deep Unfolding-Based Adaptive Filtering

Adaptive filtering is at the core of countless signal processing systems, in applications ranging from echo and noise cancellation, equalization, system identification, and adaptive beamforming, to tracking, prediction and control. Deep unfolding present a unique opportunity for legacy adaptive filtering algorithms, including least mean squares (LMS), normalized least mean squares (NLMS), recursive least squares (RLS), and Kalman filters, to become even more cost-efficient and to show enhanced performance.
Deep Unfolding-Based Adaptive Filtering

The training for DUNs removes the need for separate processes or sub-systems dedicated to approximating signal statistics, such as a signal covariance matrix. Moreover, the possibility of real-time training provides for updating the filter coefficients required to be changed (due to changes in the input or the environment).
Gradient-Descent or Stochastic-Gradient Algorithms: The well-known gradient-descent (or ascent) is a first-order iterative algorithm for finding the minimum (or maximum) of an optimization objective. When the gradient has to be approximated (e.g., by approximating signal statistics), the resulting algorithm is referred to as stochastic-gradient. First order methods are great candidates for deep unfolding. Particularly, a real-time training for the obtained DNNs removes the need for separate processes or sub-systems dedicated to approximating signal statistics, such as an interference covariance matrix.