Model-Based Deep Learning

Other titles in Foundations and Trends® in Signal Processing

Learning with Limited Samples: Meta-Learning and Applications to Communication Systems
Lisha Chen, Sharu Theresa Jose, Ivana Nikoloska, Sangwoo Park, Tianyi Chen and Osvaldo Simeone
ISBN: 978-1-63828-136-8

Signal Decomposition Using Masked Proximal Operators
Bennet E. Meyers and Stephen P. Boyd
ISBN: 978-1-63828-102-3

Online Component Analysis, Architectures and Applications
João B. O. Souza Filho, Lan-Da Van, Tzyy-Ping Jung and Paulo S. R. Diniz
ISBN: 978-1-63828-116-0

An Introduction to Quantum Machine Learning for Engineers
Osvaldo Simeone
ISBN: 978-1-63828-058-3

Bilevel Methods for Image Reconstruction
Caroline Crockett and Jeffrey A. Fessler
ISBN: 978-1-63828-002-6

Model-Based Deep Learning

Nir Shlezinger
Ben-Gurion University of the Negev
nirshl@bgu.ac.il

Yonina C. Eldar
Weizmann Institute of Science
yonina.eldar@weizmann.ac.il

the essence of knowledge
Boston — Delft

Foundations and Trends® in Signal Processing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

N. Shlezinger and Y. C. Eldar. *Model-Based Deep Learning*. Foundations and Trends® in Signal Processing, vol. 17, no. 4, pp. 291–416, 2023.

ISBN: 978-1-63828-265-5

Foundations and Trends® in Signal Processing
Volume 17, Issue 4, 2023

Editorial Scope

Topics

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multirate signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing
  - Classification and detection
  - Estimation and regression
  - Tree-structured methods

Information for Librarians

Foundations and Trends® in Signal Processing, 2023, Volume 17, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Contents

Model-Based Deep Learning

Nir Shlezinger[1] and Yonina C. Eldar[2]

[1] *School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel; nirshl@bgu.ac.il*
[2] *Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Israel; yonina.eldar@weizmann.ac.il*

ABSTRACT

Signal processing traditionally relies on classical statistical modeling techniques. Such model-based methods utilize mathematical formulations that represent the underlying physics, prior information and additional domain knowledge. Simple classical models are useful but sensitive to inaccuracies and may lead to poor performance when real systems display complex or dynamic behavior. More recently, deep learning approaches that use highly parametric deep neural networks (DNNs) are becoming increasingly popular. Deep learning systems do not rely on mathematical modeling, and learn their mapping from data, which allows them to operate in complex environments. However, they lack the interpretability and reliability of model-based methods, typically require large training sets to obtain good performance, and tend to be computationally complex.

Model-based signal processing methods and data-centric deep learning each have their pros and cons. These paradigms can be characterized as edges of a continuous spectrum varying in specificity and parameterization. The methodologies that lie in the middle ground of this spectrum, thus integrating model-based signal processing with deep learning, are

Nir Shlezinger and Yonina C. Eldar (2023), "Model-Based Deep Learning", Foundations and Trends® in Signal Processing: Vol. 17, No. 4, pp 291–416. DOI: 10.1561/2000000113.

referred to as *model-based deep learning*, and are the focus here.

This monograph provides a tutorial style presentation of model-based deep learning methodologies. These are families of algorithms that combine principled mathematical models with data-driven systems to benefit from the advantages of both approaches. Such model-based deep learning methods exploit both partial domain knowledge, via mathematical structures designed for specific problems, as well as learning from limited data. We accompany our presentation with running signal processing examples, in super-resolution, tracking of dynamic systems, and array processing. We show how they are expressed using the provided characterization and specialized in each of the detailed methodologies. Our aim is to facilitate the design and study of future systems at the intersection of signal processing and machine learning that incorporate the advantages of both domains. The source code of our numerical examples are available and reproducible as Python notebooks.

1

Introduction

The philosophical idea of artificial intelligence (AI), dating back to the works of McCarthy from the 1950's [40], is nowadays evolving into reality. The growth of AI is attributed to the emergence of machine learning (ML) systems, which learn their operation from data, and particularly to deep learning, which is a family of ML algorithms that utilizes neural networks as a form of brain-inspired computing [35]. Deep learning is demonstrating unprecedented success in a broad range of applications: deep neural networks (DNNs) surpass human ability in classifying images [24]; reinforcement learning allows computer programs to defeat human experts in challenging games such as Go [70] and Starcraft [74]; generative models translate text into images [50] and create images of fake people which appear indistinguishable from true ones [29]; and large language models generate sophisticated documents and textual interactions [48].

While deep learning systems rely on data to learn their operation, traditional signal processing is dominated by algorithms that are based on simple mathematical models which are hand-designed from domain knowledge. Such knowledge can come from statistical models based on measurements and understanding of the underlying physics, or from fixed deterministic representations of the particular problem at hand. These knowledge-based processing algorithms, which we refer to henceforth as *model-based methods*, carry out inference based on domain knowledge of the underlying model relating the observations at hand and the desired information. Model-based methods, which form the basis for many classical and fundamental signal processing techniques, do not rely on data to learn their mapping, though data is often used to estimate a small number of parameters. Classical statistical models rely on simplifying assumptions (e.g., linear systems, Gaussian and independent noise, etc.) that make inference tractable, understandable, and computationally efficient. Simple models frequently fail to represent nuances of high-dimensional complex data, and dynamic variations, settling with the famous observation made by statistician George E. P. Box that *"Essentially, all models are wrong, but some are useful"*. The usage of mismatched modeling tends to notably affect the performance and reliability of classical methods.

The success of deep learning in areas such as computer vision and natural language processing made it increasingly popular to adopt methodologies geared towards data for tasks traditionally tackled with model-based techniques. It is becoming common practice to replace principled task-specific decision mappings with abstract purely data-driven pipelines, trained with massive data sets. In particular, DNNs can be trained in a supervised way end-to-end to map inputs to predictions. The benefits of data-driven methods over model-based approaches are threefold:

1. Purely data-driven techniques do not rely on analytical approximations and thus can operate in scenarios where analytical models are not known. This property is key to the success of deep learning systems in computer vision and natural language processing, where accurate statistical models are typically scarce.

2. For complex systems, data-driven algorithms are able to recover features from observed data which are needed to carry out inference [6]. This is sometimes difficult to achieve analytically, even when complex models are perfectly known, e.g., when the environment is characterized by a fully-known complex simulator or a partial differential equation.

3. The main complexity in utilizing ML methods is in the training stage. In most signal processing domains, this procedure is carried out offline, i.e., prior to deployment of the device which utilizes the system. Once trained, they often implement inference at a lower delay compared with their analytical model-based counterparts [22].

Despite the aforementioned advantages of deep learning methods, they are subject to several drawbacks. These drawbacks may be limiting factors particularly for various signal processing, communications, and control applications, which are traditionally tackled via principled methods based on statistical modeling. For one, the fact that massive data sets, i.e., large number of training samples, and high computational resources are typically required to train such DNNs to learn a desirable mapping, may constitute major drawbacks. Furthermore, even using pre-trained DNNs often gives rise to notable computational burden due to their immense parameterization. This is highly relevant for hardware-limited devices, such as mobile phones, unmanned aerial vehicles, and Interent of Things systems, which are often limited in their ability to utilize highly-parametrized DNNs [10], and require adapting to dynamic conditions. Furthermore, the abstractness and extreme parameterization of DNNs results in them often being treated as black-boxes; understanding how their predictions are obtained and characterizing confidence intervals tends to be quite challenging. As a result, deep learning does not offer the interpretability, flexibility, versatility, reliability, and generalization capabilities of model-based methods [42].

The limitations associated with model-based methods and conventional deep learning systems gave rise to a multitude of techniques for combining model-based processing and ML, aiming to benefit from the best of both approaches. These methods are typically application-driven, and are thus designed and studied in light of a specific task. For example, the combination of DNNs and model-based sparse recovery algorithms was shown to facilitate sparse recovery [22], [47] as well as enable compressed sensing beyond the domain of sparse signals [7], [77]; Deep learning was used to empower regularized optimization methods [3],

[16], [19], while model-based optimization contributed to the design and training of DNNs for such tasks [1], [39], [78]; Digital communication receivers used DNNs to learn to carry out and enhance symbol detection and decoding algorithms in a data-driven manner [43], [63], [65], while symbol recovery methods enabled the design of model-aware deep receivers [23], [30], [57]. The proliferation of hybrid model-based/data-driven systems, each designed for a unique task, motivates establishing a concrete systematic framework for combining domain knowledge in the form of model-based methods and deep learning, which is the focus here.

In this monograph we present strategies for designing algorithms that combine model-based methods with data-driven deep learning techniques. While classic model-based inference and deep learning are typically considered to be distinct disciplines, we view them as edges of a continuum varying in specificity and parameterization. We build upon this characterization to provide a tutorial-style presentation of the main methodologies which lie in the middle ground of this spectrum, and combine model-based optimization with ML. This hybrid paradigm, which we coin *model-based deep learning*, is relevant to a multitude of research domains where one has access to some level of reliable mathematical modelling. While the presentation here is application-invariant, it is geared towards families of problems typically studied in the signal processing literature. This is reflected in our running examples, which correspond to three common signal processing tasks of compressed signal recovery, tracking of dynamic systems, and direction-of-arrival (DoA) estimation in array processing. These running examples are repeatedly specialized throughout the monograph for each surveyed methodology, facilitating the comparison between the considered approaches.

We begin by providing a unified characterization for inference and decision making algorithms in Section 2. There, we discuss different types of inference rules, present the running examples, and discuss the main pillars of designing inference rules, which we identify as selecting their type, setting the objective, and their evaluation procedure. Then, we show how classical model-based optimization as well as data-centric deep learning are obtained as special instances of this unified characterization in Sections 3 and 4, respectively. We there also review relevant basics that are core to the design of many model-based deep learning systems, including fundamentals in convex optimization (for model-based methods) and in neural networks (for deep learning). We identify the components dictating the distinction between model-based and data-driven methodologies in the formulated objectives, the corresponding decision rule types, and their associated parameters.

The main bulk of this monograph, which builds upon the fundamental aspects presented in Sections 2-4, is the review of hybrid model-based deep learning methodologies in Section 5. A core principle of model-based deep learning is to leverage data by converting classical algorithms into trainable models with varying levels of abstractness and specificity, as opposed to the more classical model-based approach where data is used to characterize the underlying model. These two rationales are highly related to the ML paradigms of generative and discriminative learning [27], [45]. Consequently, we commence this part by presenting a spectrum of decision making approaches which vary in specificity and parameterization, with model-based methods and deep learning constituting its edges, followed by a review of generative and discriminative learning. Based on these concepts, we provide a systematic categorization of model-based deep learning techniques as concrete strategies positioned along the continuous spectrum.

We categorize model-based deep learning methods into three main strategies:

1. *Learned optimization*: This approach is highly geared towards classical optimization and aims at leveraging data to fit model-based solvers. In particular, learned optimizers use automated deep learning techniques to tune parameters conventionally configured by hand.

2. *Deep unfolding*: This family of techniques converts iterative optimizers into trainable parametrized architectures. Its instances notably vary in their parameterization and abstractness based on the interplay imposed in the system design between the trainable architecture and the model-based algorithm from which it originates.

3. *DNN-aided inference*: These schemes augment model-based algorithms with trainable neural networks, encompassing a broad family of different techniques which vary in the module being replaced with a DNN.

We exemplify the considered methodologies for the aforementioned running examples via both analytical derivations as well as simulations. By doing so, we provide a systematic qualitative and quantitative comparison between representative instances of the detailed approaches for signal processing oriented scenarios. The source code used for the results presented in this monograph is available as Python Notebook scripts[1], detailed in a pedagogic fashion such that they can be presented alongside lectures, either as a dedicated graduate level course, or as part of a course on topics related to ML for signal processing.

[1]The source code and Python Notebooks can be found online at https://github.com/ShlezingerLab/MBDL_Book.

References

[1] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, 2018, pp. 394–405.

[2] A. Agrawal, S. Barratt, and S. Boyd, "Learning convex optimization models," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 8, 2021, pp. 1355–1364.

[3] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Processing Magazine*, vol. 37, no. 1, 2020, pp. 105–116.

[4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, 2002, pp. 174–188.

[5] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1983.

[6] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends in Machine Learning*, vol. 2, no. 1, 2009, pp. 1–127.

[7] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning*, JMLR, pp. 537–546, 2017.

[8] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[9] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[10] J. Chen and X. Ran, "Deep learning with edge computing: A review," *Proceedings of the IEEE*, 2019.

[11] T. Chen, X. Chen, W. Chen, H. Heaton, J. Liu, Z. Wang, and W. Yin, *Learning to optimize: A primer and a benchmark*, arXiv preprint arXiv:2103.12828, 2021.

[12] G. Choi, J. Park, N. Shlezinger, Y. C. Eldar, and N. Lee, "Split-KalmanNet: A robust model-based deep learning approach for state estimation," *IEEE Transactions on Vehicular Technology*, 2023, early access.

[13] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[14] G. Dardikman-Yoffe and Y. C. Eldar, "Learned SPARCOM: Unfolded deep super-resolution microscopy," *Optics express*, vol. 28, no. 19, 2020, pp. 27 736–27 763.

[15] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, 2004, pp. 1413–1457.

[16] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, 2018, pp. 2305–2318.

[17] T. Gafni, N. Shlezinger, K. Cohen, Y. C. Eldar, and H. V. Poor, "Federated learning: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 39, no. 3, 2022, pp. 14–41.

[18] V. Garcia Satorras, Z. Akata, and M. Welling, "Combining generative and discriminative models for hybrid inference," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[19] D. Gilton, G. Ongie, and R. Willett, "Neumann networks for inverse problems in imaging," *IEEE Transactions on Computational Imaging*, vol. 6, 2019, pp. 328–343.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, pp. 2672–2680, 2014.

[22] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *International Conference on Machine Learning*, pp. 399–406, 2010.

[23] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Transactions on Signal Processing*, vol. 68, 2020, pp. 1702–1715.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, 2020, pp. 6840–6851.

[26] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: A separable solution," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2893–2894, 1988.

[27] T. Jebara, *Machine learning: discriminative and generative*, vol. 755. Springer Science & Business Media, 2012.

[28] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, 1960, pp. 35–45.

[29] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

[30] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, 2020, pp. 5635–5648.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[33] I. Klein, G. Revach, N. Shlezinger, J. E. Mehr, R. J. van Sloun, and Y. C. Eldar, "Uncertainty in data-driven Kalman filtering for partially known state-space models," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3194–3198, 2022.

[34] O. Lavi and N. Shlezinger, "Learn to rapidly and robustly optimize hybrid precoding," *arXiv preprint arXiv:2301.00369*, 2023.

[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015.

[36] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, 2020, pp. 50–60.

[37] Y. Li, M. Tofighi, J. Geng, V. Monga, and Y. C. Eldar, "Efficient and interpretable deep blind image deblurring via algorithm unrolling," *IEEE Transactions on Computational Imaging*, vol. 6, 2020, pp. 666–681.

[38] B. Luijten, R. Cohen, F. J. De Bruijn, H. A. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. Van Sloun, "Adaptive ultrasound beamforming using deep learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, 2020, pp. 3967–3978.

[39] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by RED," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

[40] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955," *AI magazine*, vol. 27, no. 4, 2006, pp. 12–12.

[41] J. P. Merkofer, G. Revach, N. Shlezinger, and R. J. van Sloun, "Deep augmented MUSIC algorithm for data-driven DoA estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 3598–3602, 2022.

[42] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, 2021, pp. 18–44.

[43] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, 2018, pp. 119–131.

[44] A. Nakai-Kasai and T. Wadayama, "Deep unfolding-based weighted averaging for federated learning under heterogeneous environments," *arXiv preprint arXiv:2212.12191*, 2022.

[45] A. Ng and M. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes," *Advances in neural information processing systems*, vol. 14, 2001.

[46] Y. Noah and N. Shlezinger, "Distributed ADMM with limited communications via deep unfolding," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[47] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, 2020, pp. 39–56.

[48] OpenAI, "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[49] W. Pu, Y. C. Eldar, and M. R. Rodrigues, "Optimization guarantees for ISTA and ADMM based unfolded networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8687–8691, 2022.

[50] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, 2022.

[51] T. Raviv, S. Park, O. Simeone, Y. C. Eldar, and N. Shlezinger, "Online meta-learning for hybrid model-based deep receivers," *IEEE Transactions on Wireless Communications*, 2023, early access.

[52] G. Revach, N. Shlezinger, T. Locher, X. Ni, R. J. van Sloun, and Y. C. Eldar, "Unsupervised learned Kalman filtering," in *European Signal Processing Conference (EUSIPCO)*, 2022.

[53] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. van Sloun, and Y. C. Eldar, "KalmanNet: Neural network aided Kalman filtering for partially known dynamics," *IEEE Transactions on Signal Processing*, vol. 70, 2022, pp. 1532–1547.

[54] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International Conference on Machine Learning*, PMLR, pp. 1530–1538, 2015.

[55] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[56] Y. B. Sahel, J. P. Bryan, B. Cleary, S. L. Farhi, and Y. C. Eldar, "Deep unrolled recovery in sparse biological imaging: Achieving fast, accurate results," *IEEE Signal Processing Magazine*, vol. 39, no. 2, 2022, pp. 45–57.

[57] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, 2019, pp. 2554–2564.

[58] J. Scarlett, R. Heckel, M. R. Rodrigues, P. Hand, and Y. C. Eldar, "Theoretical perspectives on deep learning methods in inverse problems," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 3, 2022, pp. 433–453.

[59] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, 1986, pp. 276–280.

[60] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2014.

[61] N. Shlezinger and I. V. Bajic, "Collaborative inference for AI-empowered IoT devices," *IEEE Internet of Things Magazine*, vol. 5, no. 4, 2022, pp. 92–98.

[62] N. Shlezinger, Y. C. Eldar, and S. P. Boyd, "Model-based deep learning: On the intersection of deep learning and optimization," *IEEE Access*, vol. 10, 2022, pp. 115 384–115 398.

[63] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "ViterbiNet: A deep learning based Viterbi algorithm for symbol detection," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, 2020, pp. 3319–3331.

[64] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Learned factor graphs for inference from stationary time sequences," *IEEE Transactions on Signal Processing*, vol. 70, 2021, pp. 366–380.

[65] N. Shlezinger, R. Fu, and Y. C. Eldar, "DeepSIC: Deep soft interference cancellation for multiuser MIMO detection," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, 2021, pp. 1349–1362.

[66] N. Shlezinger and T. Routtenberg, "Discriminative and generative learning for linear estimation of random signals [lecture notes]," *IEEE Signal Processing Magazine*, 2023.

[67] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proceedings of the IEEE*, vol. 111, no. 5, 2023, pp. 465–499.

[68] D. H. Shmuel, J. P. Merkofer, G. Revach, R. J. van Sloun, and N. Shlezinger, "Deep root MUSIC algorithm for data-driven DoA estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[69] A. Shultzman, E. Azar, M. R. Rodrigues, and Y. C. Eldar, "Generalization and estimation error bounds for model-based neural networks," in *The International Conference on Learning Representations (ICLR)*, 2023.

[70] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, 2017, pp. 354–359.

[71] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R. J. van Sloun, and Y. C. Eldar, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, 2019, pp. 1051–1063.

[72] I. Sutskever, *Training recurrent neural networks*. University of Toronto, 2013.

[73] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*, 2nd ed. Elsevier Science & Technology, 2020.

[74] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, 2019, pp. 350–354.

[75] X. Wei, H. van Gorp, L. Gonzalez-Carabarin, D. Freedman, Y. C. Eldar, and R. J. van Sloun, "Deep unfolding with normalizing flow priors for inverse problems," *IEEE Transactions on Signal Processing*, vol. 70, 2022, pp. 2962–2971.

[76] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*. Cambridge University Press, 2011.

[77] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A deep learning approach for image compressive sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, 2018, pp. 521–538.

[78] K. Zhang, L. V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3217–3226, 2020.