Deep unfolding based hyper-parameter optimisation for self-interference cancellation in LTE-A/5G-transceivers

C. Motz, T. Paireder, and M. Huemer
Christian Doppler Laboratory for Digitally Assisted RF Transceivers for Future Mobile Communications, Institute of Signal Processing, Johannes Kepler University Linz, Linz, Austria
✉ Email: christian.motz@jku.at

Deep unfolding is a very promising concept that allows to combine the advantages of traditional estimation techniques, such as adaptive filters, and machine learning approaches, like artificial neural networks (ANNs). These characteristics hold equally for LTE-A and 5G/wideband signals [6]. The simplification of the complex structure of the loss function as an analytic minimisation is typically infeasible and numerical methods, for instance the simplex or back-propagation algorithms, have to be applied. To limit the complexity of the minimisation for large $N$, it is beneficial to keep the hyper-parameters constant for certain iteration intervals, i.e.

$$
\hat{\rho}_n, i.e., \ n \in [m_{i-1}, m_i), \ m_i = 1, \ldots, M, \ i = 1, \ldots, P.
$$

In the following, we focus on a challenging linear noise cancellation scenario relevant in long-term evolution advanced (LTE-A)/5G mobile communications. The mitigation of self-interference in frequency-division duplex (FDD) transceivers requires fast adaptation and sufficient steady-state performance, while operating on strongly correlated signals [6]. These characteristics hold equally for LTE-A and 5G due to the high similarity of the physical layer. We apply the deep unfolding concept to this problem and highlight the importance of a suitable loss function and training set. Incorporating application-specific knowledge, like signal characteristics and parameter uncertainties, enables substantial performance improvements for normalised LMS (N-LMS), TD-LMS and variable step-size LMS (VSS-LMS) algorithms. Moreover, we provide a reasoning for an uninuitive parameter choice for the RLS algorithm.

Deep Unfolding for Adaptive Filters: An AF algorithm is fully described by its inference equation $\hat{y}_n = f_{inf}(x_n, w_{\cdots}; \rho)$ and its coefficient update $w_n = f_{up}(x_n, y_n, w_{\cdots}; \rho)$ with the input vector $x_n \in \mathbb{C}$ and the desired value $y_n \in \mathbb{C}$ at the general iteration index $n$. Unlike classical filtering algorithms, where $x_n$ is a delay-line vector, we do not place any restrictions on the input. Moreover, the formulation explicitly includes algorithms like truncated Voletti or spline AFs that are suitable for nonlinear estimation problems. Both, the inference and the update, might depend on a number of hyper-parameters, combined into the vector $\rho \in \mathbb{R}^D$. The deep unfolding concept aims to optimise $\rho_n$ with respect to a suitable loss function. A possible choice is the weighted sum of the squared estimation errors $e_n = y_n - \hat{y}_n$, given by

$$
L(\rho_1, \ldots, \rho_D) = \sum_{n=1}^N \rho_n (e_n(\rho_1, \ldots, \rho_D))^2.
$$

The common approximation of the expectation is the sample mean of different realisations $r$:

$$
\hat{L}(\rho_1, \ldots, \rho_D) = \frac{1}{N} \sum_{r=1}^N L(r(\rho_1, \ldots, \rho_D)) \approx \hat{L}(\rho_1, \ldots, \rho_D).
$$

In the following, the superscript $(r)$ indicates a specific realisation. The described structure is visualised in Figure 1 for a single realisation, highlighting the similarity to a classical ANN. It is important to note that the coefficient initialisation $w_0$, the error weighting $\tau_n$ and the hyper-parameters $\rho_n$ are independent of $r$. Due to the complex structure of the loss function an analytic minimisation is typically infeasible and numerical methods, for instance the simplex or back-propagation algorithms, have to be applied. To limit the complexity of the minimisation for large $N$, it is beneficial to keep the hyper-parameters constant for certain iteration intervals, i.e. $\hat{\rho}_n, i.e., n \in [m_{i-1}, m_i), \ m_i = 1, \ldots, M, \ i = 1, \ldots, P$.

Note that the interval counts $M$, and lengths $m_i = m_{i-1} - 1$ can be chosen individually for all $P$ hyper-parameters. The $\hat{\rho}_n, i.e.,$ are the new variables to be optimised.

The full potential of the deep unfolding concept is obtained when tuning AFs for hardware implementations. By replacing the analytic inference and update functions with their actual implementations, any numerical inaccuracies, e.g. caused by rounding errors in fixed-point arithmetic, are included in the optimisation. Combined with a sufficiently high number of realisations, the data-driven tuning helps to avoid instabilities, too.

Application to Modulated Spur Cancellation: In the following, we investigate a particular linear estimation problem to show that, under challenging conditions, substantial improvements are possible even in such a widely covered field. The cancellation of modulated spur interferences, a class of self-interferences, is relevant in LTE-A/5G transceivers. The main issues are high power differences between simultaneously occurring transmit (Tx) and receive (Rx) signals combined with an insufficient Tx-Rx isolation and receiver non-idealities. Consequently, the Tx signal leaks into the receiver, where it can create a modulated spur interference that deteriorates the wanted Rx signal. For further details on the effect we refer to [7]. In short, the resulting AF problem resembles the well-known noise cancellation application. Though, due to tough signal characteristics, this problem is challenging for low-cost AFs which are preferred to enable a real-time implementation. First, the signals are typically narrow-band in order to enable resource sharing among different users. Second, the symbol-wise nature of the involved signals leads to non-stationarity at these symbol boundaries [6]. The simplified signal model (by neglecting a known frequency offset) in the Rx baseband is

$$
Y_n = y_{inf,n} + y_{Rx,n} + \eta_n, \quad Y_{inf,n} = \hat{h}_{Tx,n} * x_n
$$

Fig. 1 Deep unfolding structure for hyper-parameter optimisation of AFs
with the interference $y_{\text{int},n}$, the signal-of-interest $y_{\text{Rx},n}$ and additive noise $\eta_n$. $h_{\text{TX},n}$ is the unknown impulse response of the leakage path and $x_n$ are the Tx baseband samples. In order to avoid degradation of the signal-of-interest, we target to cancel the interference by estimating the leakage path by an AF. Note that both the additive noise and the signal-of-interest act as noise for the cancellation process. We employ deep unfolding with two loss functions to optimise the performance of different AF algorithms and compare them in simulations. The setup follows the considerations of [7]. For the Tx we use an LTE-10 signal with an allocated bandwidth of 2.16 MHz (12 out of 50 resource blocks). The wanted Rx uses an LTE-10 downlink signal with full bandwidth and a symbol and $\tau_n$ is the Tx baseband samples. In order to avoid degradation of the interference levels, although all values are smaller compared to $\mu_{\text{NLMS}}$. The power estimation is only slightly affected by the noise level, since it operates on the exact input sequence $u_n$.

A crucial metric for the modulated spur application is the time- and ensemble-averaged signal-to-interference-plus-noise ratio (SNIR) after cancellation, given by

$$\text{SNIR} = \frac{1}{\tau} \sum_{n=1100}^{\tau} \frac{E[|y_{\text{Rx},n}|^2]}{E[|y_{\text{int},n}-y_{\text{int},n}+\eta_n|^2]}. \quad (9)$$

Note that the summation starts at $n = 1100$ to exclude the first symbol, which causes the SNIR to mainly represent the steady-state performance.

In Figure 3, we compare the SNIR of the optimised N-LMS and TD-LMS algorithms for both error weightings. Dotted vertical lines indicate the ICN values, for which the hyper-parameters of the AFs were optimised. Each grey rectangle indicates an ICN range for which a learned hyper-parameter has been used. This accounts for the fact that the exact ICN is usually unknown. While the difference between the error weightings is small for the N-LMS, in case of the TD-LMS, the lower $\tau_{2,n}$ for $n < 1100$ lead to a smaller $\mu_{\text{TD}}$, which enables a more accurate adaptation. Thus, the steady-state performance improves substantially with an SNIR gain of up to 2 dB.

**Steady-State Performance:** Based on the previous results, we solely continue with the superior error weighting $\tau_{2,n}$ and examine two additional AFs, the VSS-LMS and the RLS. The VSS-LMS [4] improves the adaptation rate of the N-LMS by employing an iteration-dependent step-size $\mu_{\text{VSS},n}$. In the standard update Equation (6). The inference Equation (5) coincides with the N-LMS. Many strategies have been proposed for optimal step-size decay, often relying on a vanishing estimation error. In case of modulated spur cancellation, the error signal contains the signal-of-interest, and thus the prerequisite of many step-size selection strategies is not fulfilled. Though, the deep unfolding concept can still be applied in a straightforward manner. In order to simplify the hyper-parameter optimisation, we allow $\mu_{\text{VSS},n}$ to change only at predefined iterations $\tau_n$ and keep it constant after the last adjustment. In total, we split the 8 intervals of increasing length and hence have 8 hyper-parameters to train. We choose an initial value of $\mu_{\text{VSS},n} = 0.1$ for all segments.

The interval lengths and the optimal step-size sequences for three interference power levels are visualised in Figure 4. Besides the clear dependence of the step-size sequences on the ICN, all curves exhibit a fast decay to very small values, which enables good steady-state performance. For the highest ICN, the step-size in one segment is very close to 0 dB.
the stability bound of 1, resulting in an extremely fast adaptation within the first samples.

Unlike the LMS variants, the RLS [3] does not require time-dependent hyper-parameters. When using the exponentially-weighted variant, the two major parameters of the RLS are the forgetting factor $\lambda$ and the initialisation factor $\alpha$ of the estimated inverse auto-covariance matrix. We use common values for initialisation, which implies a large value of $\alpha$. However, the optimal results in Table 1 indicate that an unexpectedly low $\alpha$ is superior. This result is mainly related to the adaptation behaviour as discussed in the next section.

In Figure 5, the best cancellation results of the N-LMS and TD-LMS are compared to the VSS-LMS and the RLS. Clearly, the RLS outperforms all other algorithms by almost restoring the receiver SNR of 10 dB, independent of the ICN. However, when considering their complexity, the optimised TD-LMS and even the VSS-LMS feature remarkable improvements over the standard N-LMS. Tuning the hyper-parameters manually can hardly yield a comparable cancellation performance, especially for a high parameter count.

**Adaptation Behaviour:** Besides the steady-state performance, a fast adaptation is crucial for self-interference cancellation due to fast changing system parameters especially in 5G. A universally applicable metric for the adaptation performance is the normalised mean square error (NMSE) defined as

$$\text{NMSE}_n = \frac{E[(y_{\text{int},n} - \hat{y}_{\text{int},n})^2]}{E[|y_{\text{int},n}|^2]}.$$  \hfill (10)

In Figure 6 we compare the NMSE of all covered AFs for ICN of 16 dB. The N-LMS and the TD-LMS show a comparable initial NMSE decay, but the TD variant features a steady-state improvement of about 4 dB. In the early adaptation performance, the optimised VSS-LMS can even compete with the RLS. Both reach an NMSE of $-15$ dB in less than 30 samples. At the end of the first symbol the NMSE values are as follows: $-18.8$ dB for the N-LMS, $-21.8$ dB for the TD-LMS, $-23.3$ dB for the VSS-LMS and $-35.9$ dB for the RLS. All curves exhibit regular spikes at the boundaries of the individual LTE-10 symbols, which is caused by the non-stationarity mentioned previously [6].

As we already mentioned above, the low optimal values for $\alpha$, exemplarily given in Table 1, are contradictory to the basic literature on the RLS. To analyse the effect, we compare the initial NMSE decay for $\alpha = 10^5$ and $\alpha = 1.65$ for an ICN of $16$ dB. The corresponding curves are plotted in Figure 7. A large $\alpha$ causes an overshoot of the estimation error within the first 25 samples, whereas a small $\alpha$ leads to a monotonic decrease of the error. Hence, with the optimised $\alpha$, only a minimum number of Rx samples is considerably impaired by the interference.

**Conclusion:** We demonstrated that deep unfolding can be utilised to improve the performance of AFs in challenging applications. In particular, we optimised the relevant hyper-parameters for the modulated spur self-interference cancellation problem in FDD radio frequency (RF) transceivers. We discussed the choice of the loss function and showed its influence on the steady-state and adaptation behaviour. The deep unfolding concept yields remarkable performance gains without increasing the online complexity of the AFs. Notably, it enables a straight-forward tuning of variable step-size sequences without employing heuristics or infeasible manual optimisation.

**Acknowledgment:** The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association are gratefully acknowledged.

© 2021 The Authors. *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 16 February 2021 Accepted: 7 May 2021
doi: 10.1049/ell2.12230

**References**

1 Balatsoukas-Stimming, A. et al.: Deep unfolding for communications systems: A survey and some new directions. In: 2019 IEEE International Workshop on Signal Processing Systems, pp. 266–271. IEEE, Piscataway, NJ (2019)

2 Chen, M., et al.: Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Commun. Surv. Tutor.* 21(4), 3039–3071 (2019)

3 Diniz, P. S.: *Adaptive Filtering: Algorithms and Practical Implementation*, Second Edition, Kluwer Academic Publishers, Dordrecht (2002)

4 Harris, R., et al.: A variable step (VS) adaptive filter algorithm. *IEEE Trans. Acoust. Speech Signal Process.* 34(2), 309–316 (1986)

5 Hershey, J. R., et al.: Deep unfolding: Model-based inspiration of novel deep architectures. arXiv:1409.2574, (2014)

6 Motz, C. et al.: Improving digital interference cancellation in LTE-A/5G transceivers by statistical modeling. Asilomar Conf. Signals, Syst., Comput., (2020)

7 Motz, C. et al.: Modulated spur interference cancellation for LTE-A/5G transceivers: A system level analysis. In: 2020 IEEE 91st Vehicular Technol. Conf., pp. 1–6. IEEE, Piscataway, NJ (2020)