Neural Networks For Nonlinear Fourier Spectrum Computation

Egor Sedov[1, 2], Pedro J. Freire[1], Igor Chekhovskoy[2], Sergei Turitsyn[1], Jaroslaw Prilepsky[1]

[1] Aston Institute of Photonic Technologies, Aston University, Birmingham, B4 7ET, UK,
[2] Novosibirsk State University, Novosibirsk, 630090, Russia, e.sedov@g.nsu.ru

2021 European Conference on Optical Communication (ECOC) | 978-1-6654-3868-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ECOC52684.2021.9605844

Abstract *We demonstrate that neural networks can outperform conventional numerical nonlinear Fourier transform algorithms for processing the noise-corrupted optical signal. Applying the Bayesian hyper-parameters optimisation, we design the architecture of neural networks capable to compute nonlinear signal spectrum at low SNR more accurately than conventional algorithms.*

Introduction

The nonlinear Fourier transform (NFT) is an unconventional approach to manage nonlinear signal distortions in fibre channels[1],[2]. The direct NFT corresponds to presenting an optical signal on some special basis, called the nonlinear spectrum. The latter consists of discrete and continuous parts that evolve trivially in a linear manner in the communication channel governed by the nonlinear Schrödinger equation[2],[3]. It has been demonstrated that the transmission systems based on the NFT utilisation and modulation of nonlinear spectrum modes have the potential to outperform the conventional Fourier-based systems[4]–[7]. However, one of the serious challenges in the application of the NFT in high-speed optical communication systems is its capability to deal with the optical signals substantially perturbed by noise[6],[8],[9].

A number of numerical methods for the computation of nonlinear spectrum have been proposed and tested in optical transmission applications[10]–[12]. Significant progress has been achieved in reducing the complexity of the algorithms (fast NFT)[13],[14] and in improving their accuracy[15]–[17]. However, at the moment, the real-time NFT-based processing of complex waveforms is still far from being practically implementable at the hardware level[18], and the NFT systems' performance suffers from the deviations of the real systems from the idealised channel model[5],[6],[8],[9]. An attractive approach to overcome these difficulties, can be use of machine learning and, in particular, the neural networks (NN) for the NFT-based signal processing[19],[20]. In this work, we demonstrate that using a special advanced NN architecture that we implement using the Bayesian hyper-parameter optimisation for NFT, we can achieve the high denoising level in the recovered nonlinear spectrum.

Methods and results

To exemplify the method's functioning, without loss of generality we consider the signal in the form of a single WDM return-to-zero symbol represented as a sum of the independent optical carriers[21]:

$$q(t) = \sum_{k=1}^{M} C_k e^{i\omega_k t} f(t),\ 0 \le t < T, \qquad (1)$$

where M is a number of WDM channels, ω_k is a carrier frequency of the k-th channel, C_k corresponds to the digital data in k-th channel, and T is the symbol interval. $f(t)$ is a waveform of a return-to-zero carrier pulse, which in this work is taken (in the normalized form) as: $f(t) = \left[1 - \cos\left(\frac{4\pi t}{T}\right)\right]$ for $0 \le t \le \frac{T}{4}$ or $\frac{3T}{4} \le t \le T$, and $f(t) = 1$ for $\frac{T}{4} < t < \frac{3T}{4}$.

The direct NFT maps a signal $q(t)$ to the so-called nonlinear spectrum, which is acheived through the solution of the Zakharov-Shabat spectral problem[3]:

$$\begin{cases} -\partial_t \psi_1 + q(t)\psi_2 = i\lambda\psi_1, \\ \partial_t \psi_2 + q^*(t)\psi_1 = i\lambda\psi_2, \end{cases} \qquad (2)$$

where ψ_i are the auxiliary functions, ξ is the spectral parameter (the nonlinear analogue of conventional Fourier frequency), and $q(t)$ is the considered optical signal. Our goal here is to compute the continuous nonlinear spectrum $r(\xi)$ considering the solutions of 2 with special boundary conditions at the trailing signal's end, $t \to -\infty$. The details of the NFT mathematics can be found in[2].

To train the NN for computing the nonlinear spectrum, we used 94035 signals of the form Eq. (1), with C_k for each carrier randomly drawn from QPSK constellation; the number of optical channels (carriers) in (1) was 15. Then we sampled our signal at equidistant points in time, t_m,

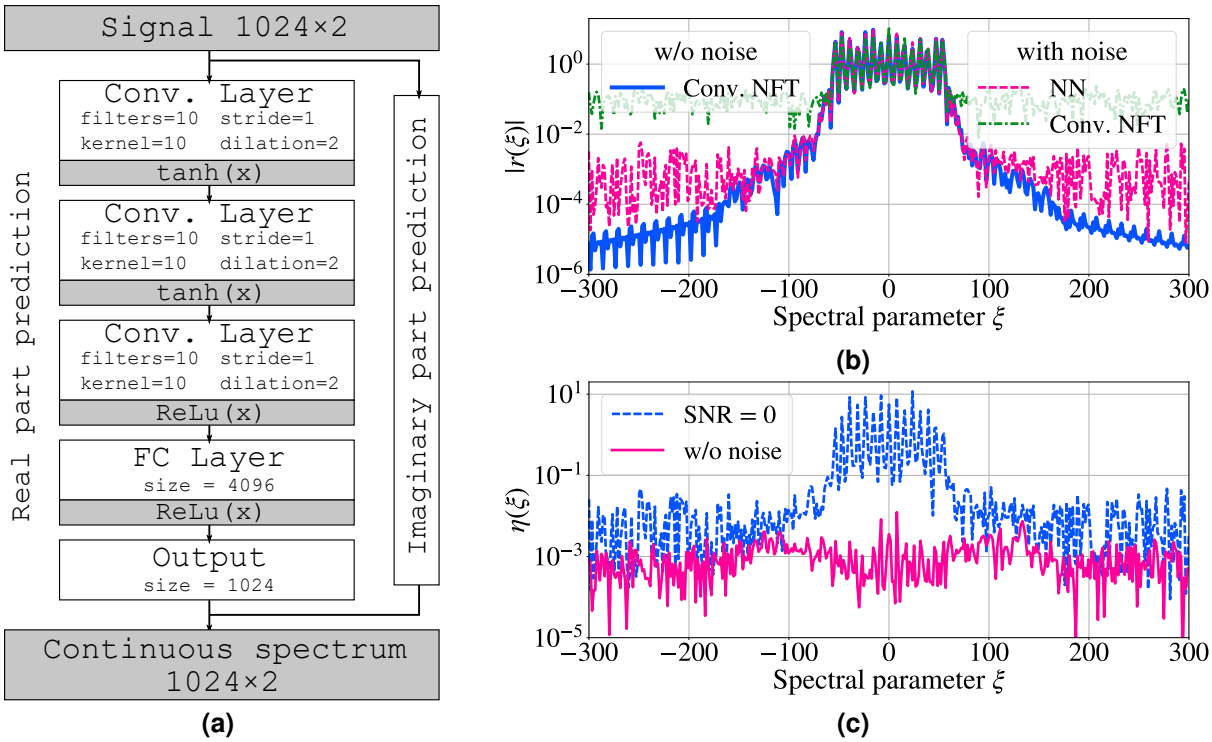

Fig. 1: (a) Neural network architecture. The parameters of each layer were determined using Bayesian optimisation. Two identical neural networks are used to predict the real and imaginary parts of the continuous spectrum samples. **(b)** Example of the amplitude of continuous nonlinear spectrum $r(\xi)$ calculated using the NN from **(a)** and with the conventional NFT method. The blue solid line shows the spectrum for the noiseless case. Green and red dashed lines show the calculated spectra for a noisy signal. **(c)** Example of metric (4) for continuous spectrum predicted by the NN from **(a)** trained on the signals without noise (red solid line) and on signals with high noise at SNR = 0 (blue dashed line).

over the segment of length T, $q(t_m) = q_m$: the number of sample points in each signal representation was $2^{10} = 1024$. The normalised symbol interval T was set to unity, so that the time step size used was $\Delta t = 2^{-10}$. For the generated discretised profile, the reflection coefficient $r(\xi)$ was identified for 1024 sample points in ξ variable, calculated using the fast numerical NFT method[14]. The parameter ξ for our computations ranged from $-\pi/(4\Delta t) \approx -804$ to $\pi/(4\Delta t) \approx 804$. Each signal in the dataset was eventually normalised so that its energy was set to $E_{\text{signal}} = 39.0$. We made sure that none of our training/testing signals contained solitons. Then, we also generated the signal sets adding uncorrelated Gaussian random value to each sample point. For further training, in addition to the set without noise, which had 84632 signals, we used 8 sets of 423160 signals (5 different noise realisations). Each set corresponds to one of the following SNR values: $\{0, 5, 10, 13, 17, 20, 25, 30\}$ dB. 9 sets of 9403 signals with the corresponding noise levels were left to validate the network performance.

We used Bayesian optimization to determine the best neural network architecture for the problem. The Bayesian optimization builds a probabilistic model of the function mapping from hyper-parameter values to the objective evaluated on a validation set[22],[23]. Fig. 1a depicts the schematic for the entire optimised NN architecture. The convolutional part consists of three layers with 10, 15 and 10 filters. Kernel sizes of the first and third convolutional layers are 10, and for the layer between them, it is 18. As noted above, we took the dilation value for each layer as one of the sought hyperparameters. The exemplary picture of how the designed NN works on one signal is given in Fig. 1b. Already from this figure, we can notice that the result produced by our NN and the nonlinear spectrum profiles obtained with conventional NFT[14] are very similar. The prediction quality of our NN is shown in Fig. 1c, which demonstrates the quality metric $\eta(\xi)$, Eq. (4), for the case of signals with (SNR = 0) and without noise.

To analyse our NN 's performance and denoising capabilities, we compare the deviations in the obtained nonlinear spectrum calculated with the NN and with the result of the conventional NFT applied to the same signal without noise (the benchmarking zero-noise level). To quantify the performance rendered by the NN with that of conventional algorithms applied to noisy signals, we use the following metric:

Tab. 1: Comparison of the NN performance against the conventional NFT in the computation of the nonlinear spectrum $r(\xi)$. The values in the cells show error value (3) for each specific pair of training and validation sets SNR. The grey cells correspond to the cases when the accuracy of the NN nonlinear spectrum restoration is lower than that of fast NFT, while the white cells correspond to the cases when the accuracy of the continuous NF spectrum predicted by the NN is higher.

			Training SNR level, dB								
		Conv. NFT	w/o noise	30	25	20	17	13	10	5	0
Validation SNR level, dB	w/o noise	0	8.39e-4	6.52e-3	9.43e-3	1.26e-2	1.61e-2	2.38e-2	3.59e-2	7.43e-2	1.42e-1
	30	6.91e-2	5.54e-2	9.56e-3	1.11e-2	1.36e-2	1.68e-2	2.42e-2	3.63e-2	7.49e-2	1.44e-1
	25	1.23e-1	9.84e-2	1.40e-2	1.39e-2	1.51e-2	1.78e-2	2.45e-2	3.63e-2	7.47e-2	1.43e-1
	20	2.21e-1	1.74e-1	2.53e-2	2.18e-2	1.97e-2	2.08e-2	2.58e-2	3.65e-2	7.40e-2	1.43e-1
	17	3.10e-1	2.41e-1	3.96e-2	3.23e-2	2.63e-2	2.53e-2	2.78e-2	3.70e-2	7.31e-2	1.42e-1
	13	4.89e-1	3.66e-1	7.74e-2	6.12e-2	4.53e-2	3.97e-2	3.54e-2	3.98e-2	7.06e-2	1.38e-1
	10	6.78e-1	4.88e-1	1.29e-1	1.03e-1	7.36e-2	6.23e-2	5.12e-2	4.85e-2	6.87e-2	1.33e-1
	5	1.16e+0	7.26e-1	2.73e-1	2.31e-1	1.72e-1	1.43e-1	1.15e-1	9.93e-2	7.98e-2	1.17e-1
	0	2.00e+0	9.48e-1	4.79e-1	4.37e-1	3.60e-1	3.12e-1	2.59e-1	2.29e-1	1.74e-1	1.16e-1

$$\eta = \frac{1}{S}\sum_{i=1}^{S}\langle\eta_i(\xi)\rangle_\xi, \tag{3}$$

$$\eta_i(\xi) = \frac{|\{r_{\text{predicted}}(\xi)\}_i - \{r_{\text{actual}}(\xi)\}_i|}{\langle|\{r_{\text{actual}}(\xi)\}_i|\rangle_\xi}, \tag{4}$$

where S is the total number of signals in the validation set, $\langle\cdot\rangle_\xi$ denotes the averaging over the spectral interval, $\{r_{\text{predicted}}(\xi)\}_i$ and $\{r_{\text{actual}}(\xi)\}_i$ correspond to the value of nonlinear spectrum $r(\xi)$ computed for the signal number i at point ξ (we compare the quantities for the validation data set). The label "predicted" refers to the result produced by the NN on the noisy signal, and "actual" marks the $r(\xi)$ value obtained using the conventional NFT algorithm[14] for the noiseless signal. The relative error $\eta(\xi)$ is determined at the point ξ, so we use $\langle\eta(\xi)\rangle_\xi$ to estimate the overall mean of the error for one signal, and use Eq. (3) to evaluate the error for the entire validation dataset. We stress that the metric was chosen in such a way as to take into account even the regions where the value of the spectrum is much less than one.

Table 1 shows the error values for the restoration of $r(\xi)$ coefficient of a noiseless and noisy-perturbed signals (1), by the NFT-Net architecture given in Fig. 1a. The first row in the table corresponds to the noiseless case. It is always shadowed with grey, which means that the NN cannot provide any better results than the benchmark ones rendered by the conventional fast NFT method used to generate the training data.

However, the values of the error for noise-corrupted signals reveal interesting tendencies. It follows from the table that for the low training noise level (up to 10 dB, columns three through nine), the NN error is typically lowest for the noiseless validation dataset (second row). Thus, the addition of low noise in the training dataset only degrades the NN restoration capability, even though this decrease is not significant. For the most interesting case of high noise level (i.e. accumulated in long-haul systems), the network works best for the signal sets where the SNR value is the same for the validation and training sets. In such cases, the relative error is about 8-12%, while the error for conventional NFT is at the level of 100-200%. With decreasing noise (rows from bottom to top) in the validation set, the error value remains at approximately the same level after the cell corresponding to the same training and validation SNRs. These results confirm that the presented NN architecture is capable of performing the NFT operation for complicated signals, and, in addition, it can also work as a highly efficient denoising element when the noise level becomes non-negligible.

Conclusions

We examined the implementation of the NFT by the NN with a special structure, applying the Bayesian hyperparameters optimisation to determine the best-performing network architecture. We considered here an almost unexplored case dealing with the computation of the continuous nonlinear spectrum, which is most interesting from the optical communications perspective. It was demonstrated that the NN-based processing can substantially improve the accuracy of the NFT in the presence of noise, compared to the conventional high-accuracy NFT processing method. The advantage in denoising becomes most pronounced at high noise levels, with the maximum restoration quality typically occurring when the SNR of the training data is the same as that of the validation dataset.

Acknowledgements

ES and ST are supported by the EPSRC programme grant TRANSNET, EP/R035342/1. JP and ST acknowledge the support of Leverhulme Trust project RPG-2018-063. ES acknowledges the support from the Russian Science Foundation under Grant 17-72-30006, IC acknowledges the grant of the President of the Russian Federation (MK-677.2020.9).

References

[1] M. Yousefi and F. Kschischang, "Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools", *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4312–4328, 2014.

[2] S. K. Turitsyn, J. E. Prilepsky, S. T. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko, "Nonlinear fourier transform for optical data processing and transmission: Advances and perspectives", *Optica*, vol. 4, no. 3, pp. 307–322, 2017.

[3] V. E. Zakharov and A. B. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media", *Soviet Physics JETP*, vol. 34, no. 1, p. 62, 1972.

[4] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis technique for optical links with lumped amplification", *Opt. Express*, vol. 23, pp. 8317–8328, 2015.

[5] S. Le, V. Aref, and H. Buelow, "Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit", *Nature Photonics*, vol. 11, no. 9, p. 570, 2017.

[6] S. Civelli, E. Forestieri, and M. Secondini, "Mitigating the impact of noise on nonlinear frequency division multiplexing", *Applied Sciences*, vol. 10, no. 24, p. 9099, 2020.

[7] M. Yousefi and X. Yangzhang, "Linear and nonlinear frequency-division multiplexing", *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 478–495, 2019.

[8] S. Civelli, E. Forestieri, and M. Secondini, "Why noise and dispersion may seriously hamper nonlinear frequency-division multiplexing", *IEEE Photonics Technology Letters*, vol. 29, no. 16, pp. 1332–1335, 2017.

[9] M. Pankratova, A. Vasylchenkova, S. A. Derevyanko, N. B. Chichkov, and J. E. Prilepsky, "Signal-noise interaction in optical-fiber communication systems employing nonlinear frequency-division multiplexing", *Physical Review Applied*, vol. 13, no. 5, p. 054 021, 2020.

[10] G. Boffetta and A. R. Osborne, "Computation of the direct scattering transform for the nonlinear schroedinger equation", *Journal of computational physics*, vol. 102, no. 2, pp. 252–264, 1992.

[11] S. Burtsev, R. Camassa, and I. Timofeyev, "Numerical Algorithms for the Direct Spectral Transform with Applications to Nonlinear Schrödinger Type Systems", *Journal of computational physics*, vol. 147, no. 1, pp. 166–186, 1998.

[12] A. Vasylchenkova, J. E. Prilepsky, and S. K. Turitsyn, "Contour integrals for numerical computation of discrete eigenvalues in the Zakharov-Shabat problem", *Optics Lett.*, vol. 43, no. 15, pp. 3690–3693, 2018.

[13] S. Wahls and H. V. Poor, "Introducing the fast nonlinear Fourier transform", in *International Conference on Acoustics, Speech and Signal Processing*, Vancouver: IEEE, 2013, pp. 5780–5784.

[14] S. Wahls, S. Chimmalgi, and P. Prins, "FNFT: A Software Library for Computing Nonlinear Fourier Transforms", *The Journal of Open Source Software*, vol. 3, p. 597, 23 2018, ISSN: 2475-9066.

[15] R. Mullyadzhanov and A. Gelash, "Direct scattering transform of large wave packets", *Optics Letters*, vol. 44, no. 21, p. 5298, Nov. 2019.

[16] S. Medvedev, I. Chekhovskoy, I. Vaseva, and M. Fedoruk, "Conservative multi-exponential scheme for solving the direct Zakharov-Shabat scattering problem", *Optics Letters*, vol. 45, no. 7, p. 2082, Apr. 2020.

[17] S. Medvedev, I. Vaseva, I. Chekhovskoy, and M. Fedoruk, "Exponential fourth order schemes for direct Zakharov-Shabat problem", *Optics Express*, vol. 28, no. 1, p. 20, Jan. 2020. arXiv: 1908.11725.

[18] A. Vasylchenkova, D. Salnikov, D. Karaman, O. G. Vasylchenkov, and J. E. Prilepskiy, "Fixed-point realization of fast nonlinear Fourier transform algorithm for FPGA implementation of optical data processing", in *Nonlinear Optics and Applications XII*, International Society for Optics and Photonics, vol. 11770, 2021, p. 1 177 016.

[19] R. T. Jones, S. Gaiarin, M. P. Yankov, and D. Zibar, "Time-domain neural network receiver for nonlinear frequency division multiplexed systems", *IEEE Photonics Technology Letters*, vol. 30, no. 12, pp. 1079–1082, 2018.

[20] Y. Wu, L. Xi, X. Zhang, Z. Zheng, J. Wei, S. Du, W. Zhang, and X. Zhang, "Robust neural network receiver for multiple-eigenvalue modulated nonlinear frequency division multiplexing system", *Optics Express*, vol. 28, no. 12, pp. 18 304–18 316, 2020.

[21] E. V. Sedov, A. A. Redyuk, M. P. Fedoruk, A. A. Gelash, L. L. Frumin, and S. K. Turitsyn, "Soliton content in the standard optical ofdm signal", *Opt. Lett.*, vol. 43, no. 24, pp. 5985–5988, Dec. 2018.

[22] J. Močkus, "On bayesian methods for seeking the extremum", in *Optimization techniques IFIP technical conference*, Springer, 1975, pp. 400–404.

[23] M. Pelikan, D. E. Goldberg, E. Cantú-Paz, *et al.*, "Boa: The bayesian optimization algorithm", in *Proceedings of the genetic and evolutionary computation conference GECCO-99*, Citeseer, vol. 1, 1999, pp. 525–532.