Optimization of structural complexity of ANN in Signal filtering applications

D. Vorkutov, O. Sumenkov and I. Tarabukin

Sirius University of Science and Technology

Information Theory and Optimization School
21st October — 3rd November, 2022

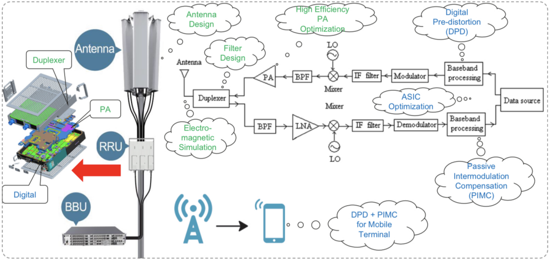
Outline

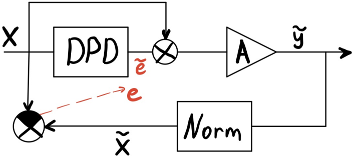
1. Introduction
2. DPD-block Cell optimization
3. 1-layer NN
4. Optuna Framework
5. 2-layer NN

Aim and targets

The aim of the project **to search for Neural Network Structure which provides well performance with reduced complexity**
The targets of the project:

- Choose optimal architecture of the Cell
- Using external optimization toolbox:
  - Search for hyper-parameters with best performance
  - Search for high performance and low complexity
- Extending the architecture with extra modules and layers





[1] Vorobyev A. Microwave Filters Design as Example of Engineering Optimization Problem, Sirius, 25.10.2022

Methodology

Initial optimization problem:

$$\min_{W \in \mathbb{C}} \quad ||DPD(W, x) - y||^2$$

where W — signal distortion function, x — input signal, $x, W \in \mathbb{C}$.

Model training process — ADAM:

Input: learning rate $\alpha > 0$, momentum terms $\beta_1, \beta_2 > 0$, parameter vector $\theta_0 \in \mathbb{R}^n$, vectors $m_0, v_0 \in \mathbb{R}^n$

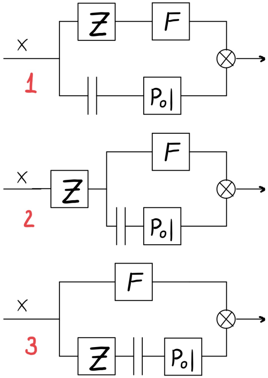
for k = 1,2, ... **do**

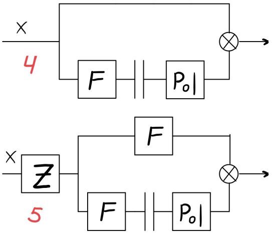
$$\text{Sample } \nabla f(x^k, \xi^r_{i=1})$$
$$\mathrm{m}_k = \beta_1 m_{k-1} + (1 - \beta_1)\nabla f(x^k, \xi^r_{i=1})$$
$$\mathrm{v}_k = \beta_2 v_{k-1} + (1 - \beta_2)\nabla f(x^k, \xi^r_{i=1})^2$$
$$_k = \frac{m_k}{1 - \beta_1^k}$$
$$_k = \frac{v_k}{1 - \beta_2^k}$$
$$\theta_k = \theta_{k-1} - \left(\frac{\alpha}{\sqrt{\hat{v}_k + \epsilon}} \hat{m}_k\right)$$

end

[2] Maslovskiy A. Gradient Descent and First-Order Momentum Optimization Algorithms, Sirius, 10.08.2020

Digital Pre-Distortion Block

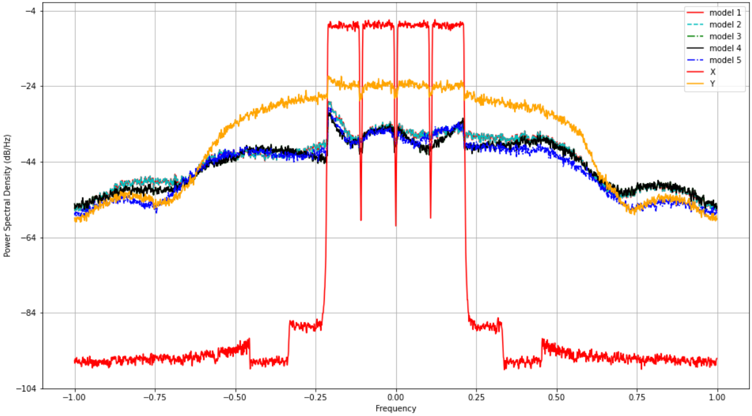




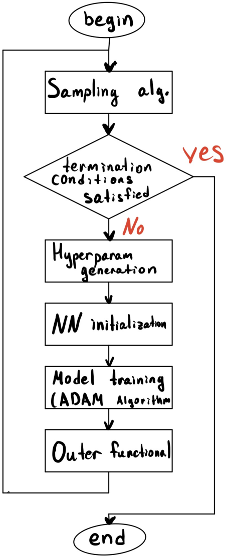
Scheme	1	2	3	4	5
score	-24.5858	-24.5806	-24.5858	-25.8865	-25.9897

F — AFIR (Arbitrary Frequency Impulse Response) block, Z — Delay block, Pol — Polynomial block, that describes the plant.

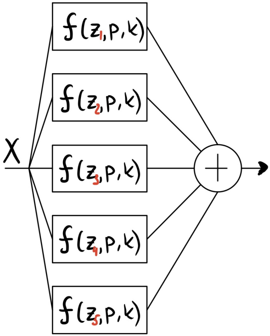
Digital Pre-Distortion Block



External cycle optimization approach for model complexity decay



DPD-block delay optimization

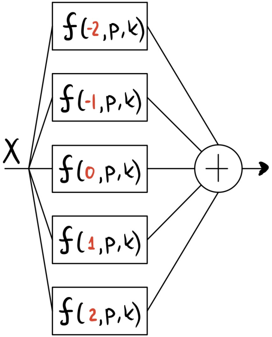


Polynomial and convolution parameters (p and k) were fixed at border values:

$$\min_{z} \sum_{i=1}^{5} f(z_i, p_{0i}, k_{0i}), \quad z \in \mathbb{Z}, \quad z \in [-10...10]$$

	z_0	z_1	z_2	z_3	z_4	Poly ord	Conv ord	Score
optuna	-3	-2	-1	2	3	4	3	-27.84
symm. princ.	-3	-2	-1	2	3	4	3	-27.84
optuna	-4	-1	0	2	3	9	9	-29.059
symm. princ.	-2	-1	0	1	2	9	9	-29.089

1-layer NN hyper-params optimization



When delay values were obtained, to reach the trade-off between performance and complexity, following functional was introduced:

$$J = \frac{com_{cur} - com_{huge}}{com_{small} - com_{huge}} + \gamma \frac{sc_{huge} - sc_{cur}}{sc_{huge} - sc_{small}}$$

where com is *complexity* $= \sum_{i=0}^{4} (p_i + k_i)$, and sc is model score.

p_0	k_0	p_1	k_1	p_2	k_2	p_3	k_3	p_4	k_4	Comp	Score
9	9	9	9	9	9	9	9	9	9	72	-33.28
9	9	8	5	6	9	9	7	6	9	77	-33.8845
9	3	7	3	4	5	5	3	4	3	46	-33.30

1-layer NN hyperparams optimization

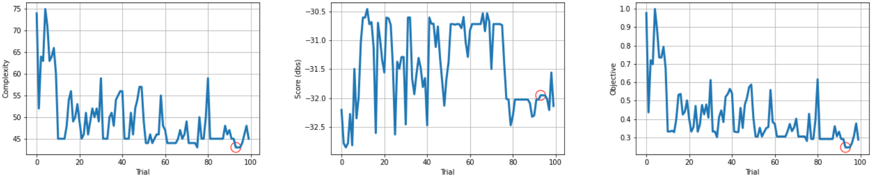


Figure: Solution of optimization problem with $\gamma = 1$

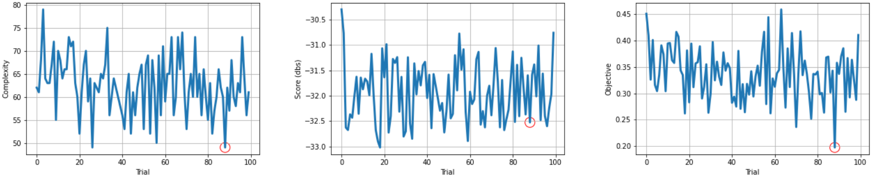


Figure: Solution of optimization problem with $\gamma = 0.3$

1-layer NN

For 1-layer NN hyperparameters (Poly order and conv. order) optimization Optuna's next algorithms were used:

	Random	TPE	GA	QMC	CMA-E
score, dbs	-31.6899	-31.8921	-31.7845	-31.7149	...
t, sec	4459.7	4781.7	4592.6	4468.8	...
score, dbs	-31.8257	-31.8582	-31.7845	-31.7149	-31.9453
t, sec	3369.0	3223.7	3508.0	3404.6	3334.7
score, dbs	-33.1475	-33.0353	-33.1095	-33.0810	-33.1353
t, sec	4049.6	4030.0	4023.5	3998.8	4271.9

Optuna Framework algorithms

For 1-layer NN hyperparameters (Poly order and conv. order) optimization Optuna's embedded algorithms were used:

	Random	TPE	NSGAII	QMC	CMA-EA
float params	+	+	+-	+	+
int params	+	+	+-	+	+
category params	+	+	+	+-	+-
Constrained opt	-	+	+	-	-
T cmpx (per tr)	$O(d)$	$O(dn \log n)$	$O(mnp)$	$O(dn)$	$O(d^3)$

TPE — Tree-Structured Parzen Estimator,

NSGAII — Nondominated Sorting Genetic Algorithm II,

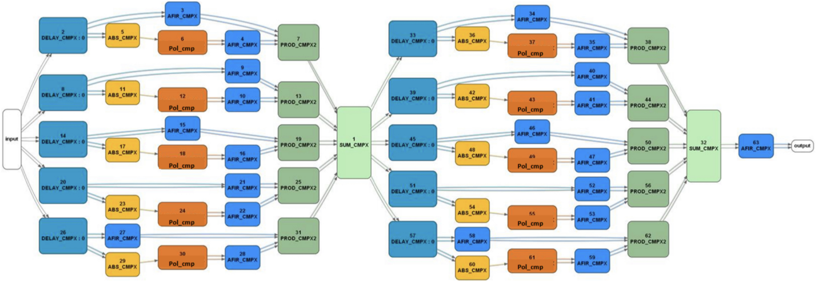
QMC — Quasi Monte-Carlo Method,

CMA-EA — Covariance Matrix Adaptation Evolution Strategy.

n — number of finished trials, d — dimension of search space, m — number of objectives

2-layers NN

To improve the performance of the NN let us connect 2 layers, so that:



Conclusion

- TO DO
- With Optuna framework trade-off between performance and model complexity was found
- Following NN was created:

NN architecture	Score	Complexity
Started single Cell	-13.8892	23
Cell №5	-25.9897	18
1-layer same blocks	-29.0898	90
1-layer Optuna huge	-33.8845	77
1-layer Optuna reduced	-32.5251	48
2-layer Optuna huge	-36.5789	270
2-layer Optuna reduced	-35.2655	155

Thank you for your attention!

D. Vorkutov, O. Sumenkov and I. Tarabukin
Sirius University of Science and Technology