We have approximated the Floating Point Operations (FLOP) count where we provide some additional results that we obtained for the non-AZ/MOAZ nomenclature, this teacher neural network can be represented as a random vector. In the linear problem, the non-linear problem is formulated according to the process discussed in [1]. A teacher neural network acts as a non-linear regression model represented as \( L(x_i) = u.\text{ReLU}(Mx_i) \) where \( M \) is a random \( 8 \times 8 \) matrix and \( u \) is a random vector. In AZ/MOAZ nomenclature, this teacher neural network can be represented as shown in Figure 1. The task of AZ/MOAZ is to rediscover this neural network by getting signals from the dataset created by using the teacher neural network.

### Table 1: The approximated number of FLOP for 65 operations used in AutoML-Zero. VSize and MSize refer to the dimensions of vectors and matrices used in the operations, respectively.

<table>
<thead>
<tr>
<th>Operators</th>
<th>Approx. FLOP</th>
<th>Operators</th>
<th>Approx. FLOP</th>
</tr>
</thead>
</table>
| ADDITIONAL DESCRIPTION OF NON-LINEAR PROBLEM

In addition to the results mentioned in the main manuscript, here we provide some additional results that we obtained for the non-linear problem. The non-linear problem is formulated according to the process discussed in [1]. A teacher neural network acts as a non-linear regression model represented as \( L(x_i) = u.\text{ReLU}(Mx_i) \) where \( M \) is a random \( 8 \times 8 \) matrix and \( u \) is a random vector. In AZ/MOAZ nomenclature, this teacher neural network can be represented as shown in Figure 1. The task of AZ/MOAZ is to rediscover this neural network by getting signals from the dataset created by using the teacher neural network.

### 2.1 Experimental Setting

To perform the experiments with MOAZ, we used the following lower and upper bounds on complexity and error: the lower bound was 0 for both objectives, whereas the upper bound was 200 for complexity and 0.4 for error. We permit all instructions used by the teacher neural network in the search space. A target error for the problem \( (\varepsilon_T) \) is defined as \( 5 \times 10^{-2} \). So, a run is considered to be successful if it could find at least one algorithm that has less than \( 5 \times 10^{-2} \) error. Both AZ and MOAZ are run 30 times with
# sX/vX/mX: scalar/vector/matrix memory
# at address X.
def Setup():
    s2 = 0
def Predict(v0):
    v2 = dot(m0, v0)
    v3 = maximum(v2, v4)
    s1 = dot(v3, v1)
def Learn(v0, s0):
    s3 = s0 - s1
    s3 = s2 * s3
    v5 = s3 * v3
    v1 = v1 + v5
    v6 = s3 * v1
    v7 = heaviside(v2, 1.0)
    v6 = v7 * v6
    m1 = outer(v6, v0)
    m0 = m0 + m1

Figure 1: Illustration of the teacher neural network used as a non-linear regression model for generating labels for non-linear regression.

different random seeds. The results and some comparisons of the performance of both frameworks are provided in Section 2.2.

2.2 Performance Comparison

The combined results for AZ and MOAZ are provided in Figure 2 and Figure 3, respectively. After combining the results for all 30 runs, the non-dominated solutions from all discovered algorithms are identified. These solutions are marked in different colors in the figures.

Figure 2: Combined results of AZ runs. Here PF refers to the Pareto Front of the combined results of AZ.

Figure 3: Combined results of MOAZ runs. Here PF refers to the Pareto Front of the combined results of AZ.

solutions are beyond 100 complexity in the case of AZ. One of the interesting applications of finding a broad distribution is platform-based designs. Depending on different platforms, users might want to use different solutions. If the computational capability of a platform is on the lower side, the users can use a low-complexity model with some trade-off in accuracy. This is not possible with AZ solutions because AZ does not take into account complexity. The solutions of MOAZ are statistically significant compared to the AZ solutions both in terms of error and complexity with $p$-value in the order of $10^{-9}$.

2.3 Discovered Algorithms

It is not possible to show all the algorithms discovered by both AZ and MOAZ. In this subsection, we are showing one representative algorithm for each of the frameworks. The AZ solution is shown in Figure 6 and the MOAZ solution is presented in Figure 7. There are not many differences between these representative algorithms, apart from the initialization in the `Setup` component. MOAZ solution initializes just the two layers in the neural network but AZ solution initializes some other vectors as well. Please note that these are discovered algorithms with more than 99% accuracy. Some
Figure 5: Swarm Plot comparison between AZ and MOAZ for complexity. In this plot, if the algorithms have equal complexity, they are placed on the same line side-by-side.

of these algorithms are very close to the teacher neural network. MOAZ also provides other solutions which are not this close to the teacher neural network but trade off some accuracy for the complexity.

Figure 6: Illustration of a working algorithm for non-linear regression discovered by AZ with a complexity of 107.

Figure 7: Illustration of a working algorithm for non-linear regression discovered by MOAZ with a complexity of 95.

REFERENCES