Report for Exercise 2 in WSI

Adam Szostek, Index Number: 331443

February 11, 2025

1 Description of the Implemented Algorithm

The implemented algorithm is the evolutionary strategy ES(1+1) with the 1/5 success rule.

Algorithm 1 ES(1+1) 1/5 Strategy

Data: $q(x)$, $\hat{x}^*$, σ, a, t_{max}
Result: $\hat{x}^*$, $\hat{o}^*$

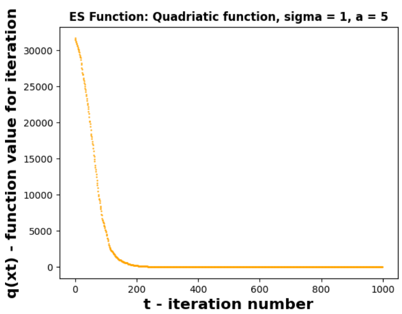
```
/* Initialization                                                  */
t ← 1 ;                                  // Current iteration number
ls ← 0 ;                 // Number of successes in the last a iterations
ô* ← q(x̂*) ;                 // Lowest objective function value for x̂*
while t ≤ t_max do
    m ← x̂* + σ · N(0,1) ;                    // New solution based on x̂*
    o_m ← q(m) ;                // Objective function value for solution m
    if o_m ≤ ô* then
        ls ← ls + 1 ;                  // Increase the number of successes
        ô* ← o_m ;               // Update the best objective function value
        x̂* ← m ;                              // Update the best solution
    end
    if t mod a = 0 then
        if ls/a > 1/5 then
            σ ← 1.22 · σ ;                   // Increase mutation strength
        end
        if ls/a < 1/5 then
            σ ← 0.82 · σ ;                   // Decrease mutation strength
        end
        ls ← 0 ;                                 // Reset success counter
    end
    t ← t + 1 ;                           // Increment iteration counter
end
```

2 Description of Planned Numerical Experiments

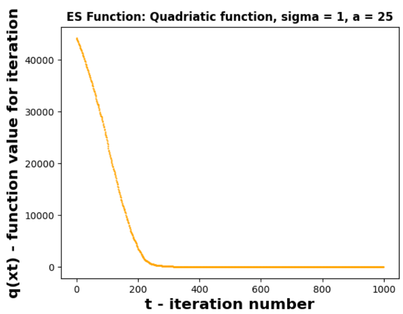
To test the effectiveness of the strategy and the influence of initial parameters on its operation, the following functions were used: a quadratic function and f3 and f7 from the CEC2017 benchmark. For each function, a convergence graph (objective function value versus iterations) was created for four different parameter variations, namely adaptation interval $a \in \{5, 25\}$ and initial mutation strength $\sigma \in \{1, 20\}$. To compensate for the stochasticity of the iterative solver, each point on the graph is the average of twenty minimization trials of the objective function. Additionally, the ES(1+1) strategy was compared with the gradient descent algorithm using the non-parametric Wilcoxon signed-rank test and convergence curves.

3 Obtained Results

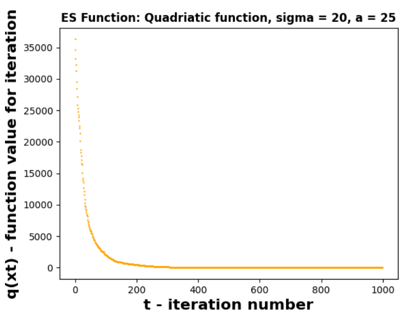
3.1 Quadratic Function



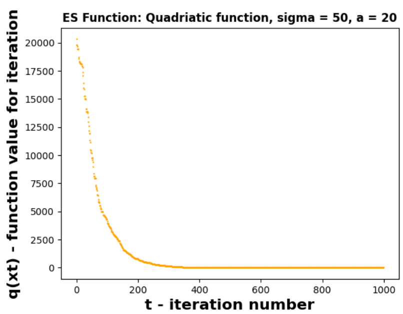
(a) a = 5, σ = 1



(b) a = 25, σ = 1

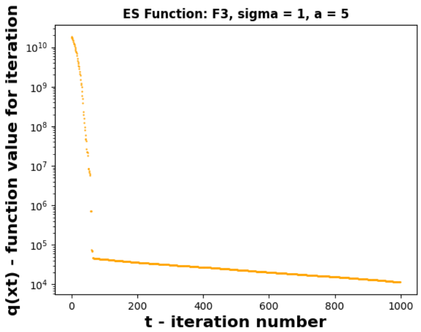


(a) a = 25, σ = 20

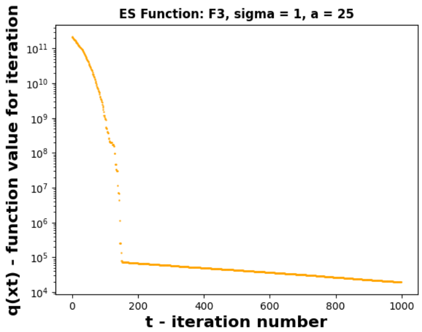


(b) a = 50, σ = 20

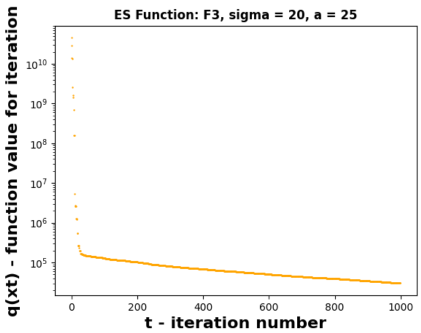
3.2 F3



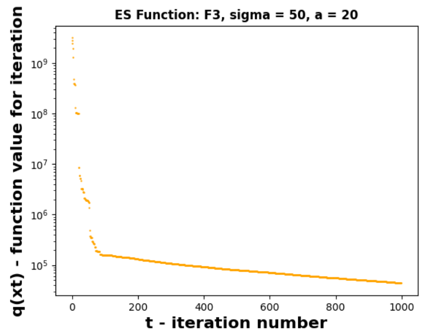
(a) a = 5, σ = 1



(b) a = 25, σ = 1

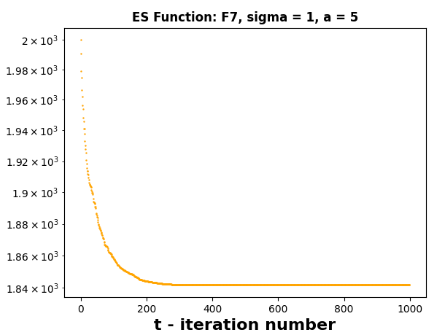


(a) a = 25, σ = 20

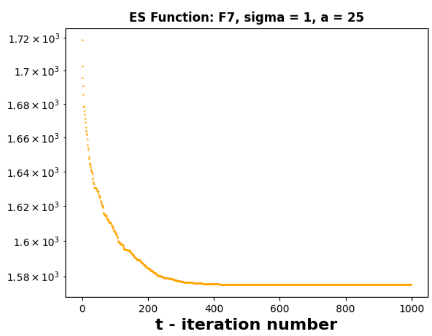


(b) a = 50, σ = 20

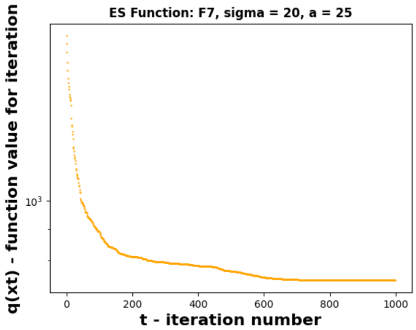
3.3 F7



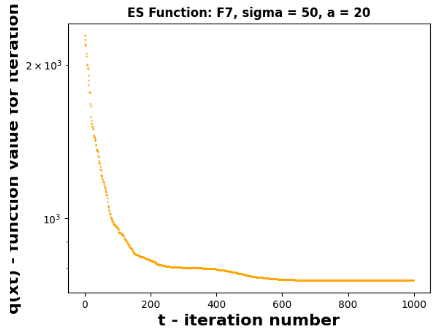
(a) a = 5, σ = 1



(b) a = 25, σ = 1



(a) a = 25, σ = 20



(b) a = 50, σ = 20

3.4 ES(1+1) 1/5 VS SGD

Below is the convergence curve comparing the ES and SGD algorithms for the F3 function from the CEC2017 benchmark. In both cases, solvers were run with parameters that, based on previous studies, were found to be optimal for the F3 function.

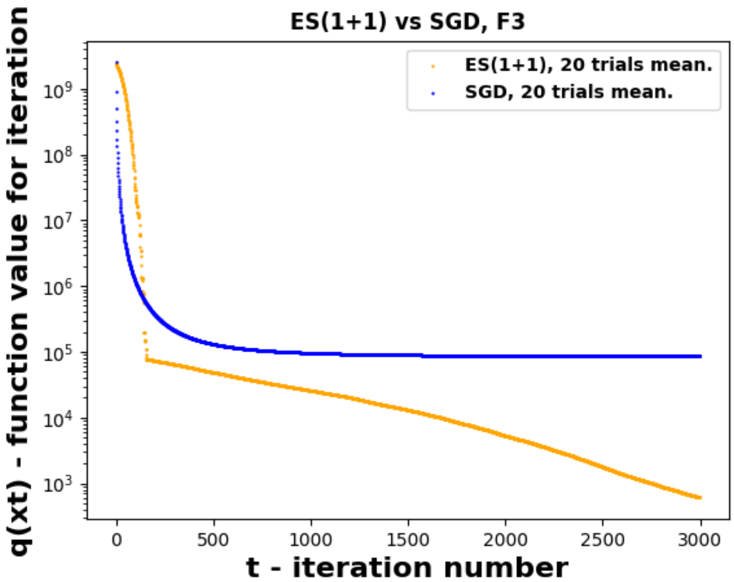


Figure 7: SGD vs ES convergence curves for the F3 function with "favorable" parameters for both algorithms

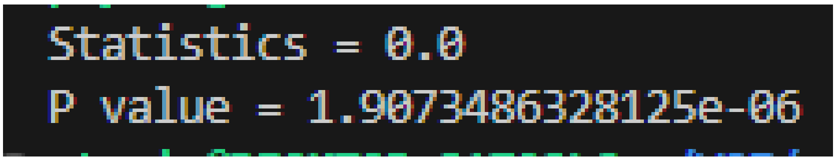


Figure 8: Results of the Wilcoxon test for 20 pairs of ES and SGD performance

4 Conclusions from the Conducted Studies

4.1 Influence of Starting Parameters on the Operation of ES(1+1)

Based on the graphs, it can be concluded that the initial mutation strength affects the initial speed at which the algorithm approaches the solution. This is evident from the density of visible points on the graphs. With a smaller initial value of σ, we observe that the algorithm initially minimizes the objective function value more slowly, whereas with a larger σ, the difference between values for successive explored points is significantly greater. After a certain number of iterations, the difference between the algorithm's performance for different initial σ values becomes negligible, as the mutation strength adapts according to the strategy's assumptions.
The adaptation interval influences the algorithm's ability to respond to changes in the solution space. When it is larger, the algorithm "focuses" on the current best solutions and more steadily strives to minimize the objective function, spending more time in the area where the current best individual is located, which favors exploitation. Conversely, a smaller adaptation interval favors exploration, as it allows for a more thorough traversal of the solution space and a faster pursuit of increasingly better individuals without focusing too long on staying in one part of the solution space. This can be observed through the "steepness" of the graphs depending on a.

4.2 ES vs SGD

Based on the convergence curve and the results of the Wilcoxon signed-rank test, it can be inferred that the ES strategy achieves better results in minimizing the objective function, i.e., its operation results in a point with a significantly lower value compared to SGD. A test statistic of 0 implies that all pairs of observations have differences indicating the advantage of one of the algorithms. A p-value on the order of 1e-6 (< 0.005) suggests that the obtained results are statistically significant. This means that the observed difference has a very low probability of being due to randomness. Observations from the convergence curves indicate that SGD initially converges more rapidly towards minimizing the objective function but later "stalls" at a point where the gradient is likely very small. In contrast, ES steadily progresses towards increasingly better minimization of the objective function value. Combining information from the Wilcoxon test with the convergence curve, it can be concluded that there is an advantage in favor of the ES strategy.