Optimization Methods Analysis

---

1. Cost History Plot: Batch Gradient Descent vs. Stochastic Gradient Descent

Observation:

- In the provided cost vs. iteration plots, stochastic gradient descent (SGD) shows a highly fluctuating cost curve compared to batch gradient descent (BGD), which follows a smoother descent.
- Despite the fluctuations, SGD achieves a lower final cost value after 50,000 iterations than BGD for the given dataset.

Theoretical Requirements:

- **BGD**: Requires computation of the gradient for the entire dataset in each iteration, which ensures a smooth and stable reduction in cost but can be computationally expensive for large datasets.
- **SGD**: Computes the gradient for a single data point per iteration, making it computationally faster per iteration but introducing fluctuations due to noise from sampling.

Practical Implications:

- **BGD**: Preferred for smaller datasets or when computational resources allow, as it is less affected by random noise.
- **SGD**: Suitable for large datasets due to faster updates, but the high variance can cause instability unless carefully tuned with learning rate decay or momentum.

---

2. Cost History Plot for Different Values of Learning Rate (α)

Observation:

- For increasing α, the cost decreases more rapidly at the start of optimization. However, overly large α values cause the cost to oscillate or even diverge.
- A balanced α ensures convergence at a reasonable rate.

Theoretical Requirements:

- A smaller α guarantees convergence but requires more iterations, increasing computational costs.
- A larger α risks overshooting the optimal solution or diverging entirely.

Practical Implications:

- Start with a moderate α and adjust using techniques like learning rate schedules (e.g., exponential decay) to optimize the convergence speed without sacrificing stability.

---

3. Line Search vs. Batch Gradient Descent

Observation:

- Line search adapts the learning rate dynamically, reducing the need for manual tuning of α.
- The cost decreases faster than BGD during early iterations due to better step size selection. However, the computational cost per iteration is higher due to the additional line search procedure.

Theoretical Requirements:

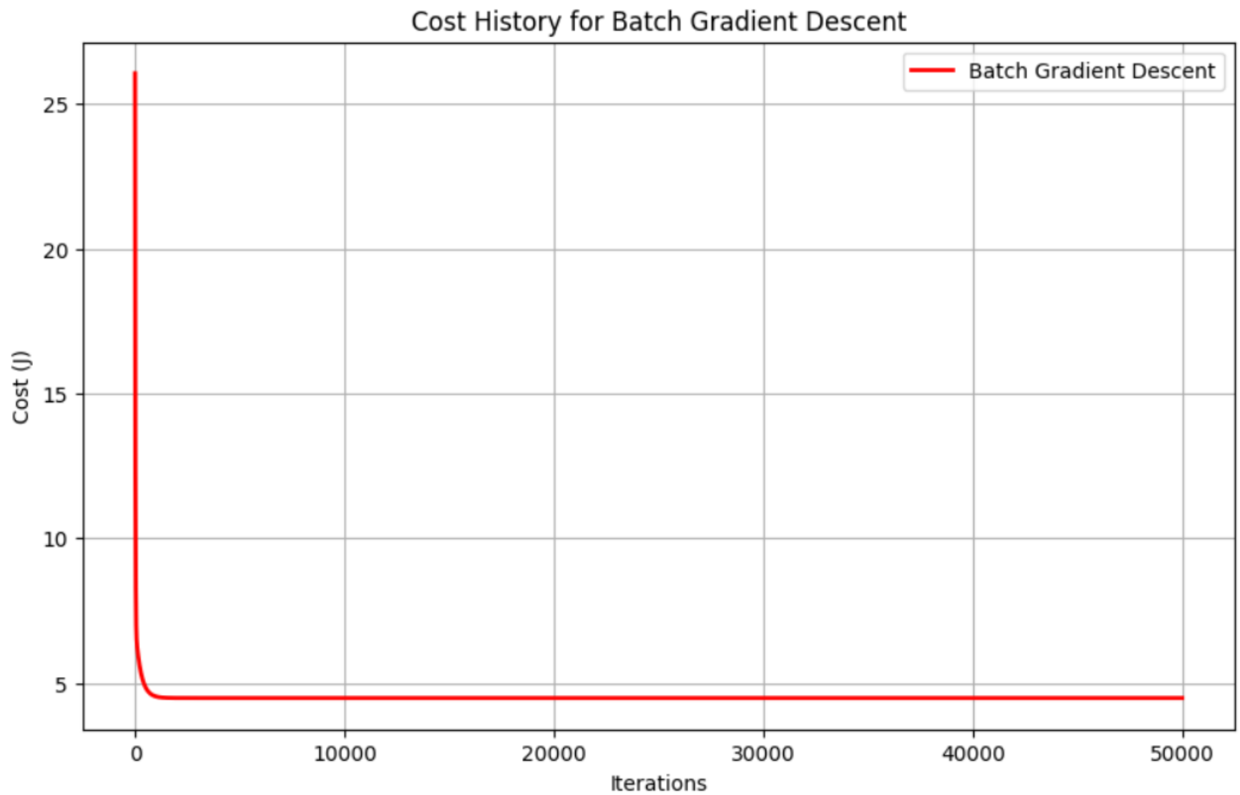
- Line search ensures that each step is optimal based on the current gradient, reducing the risk of overshooting or slow convergence.
- BGD requires a preselected α, which may not be ideal for all stages of optimization.

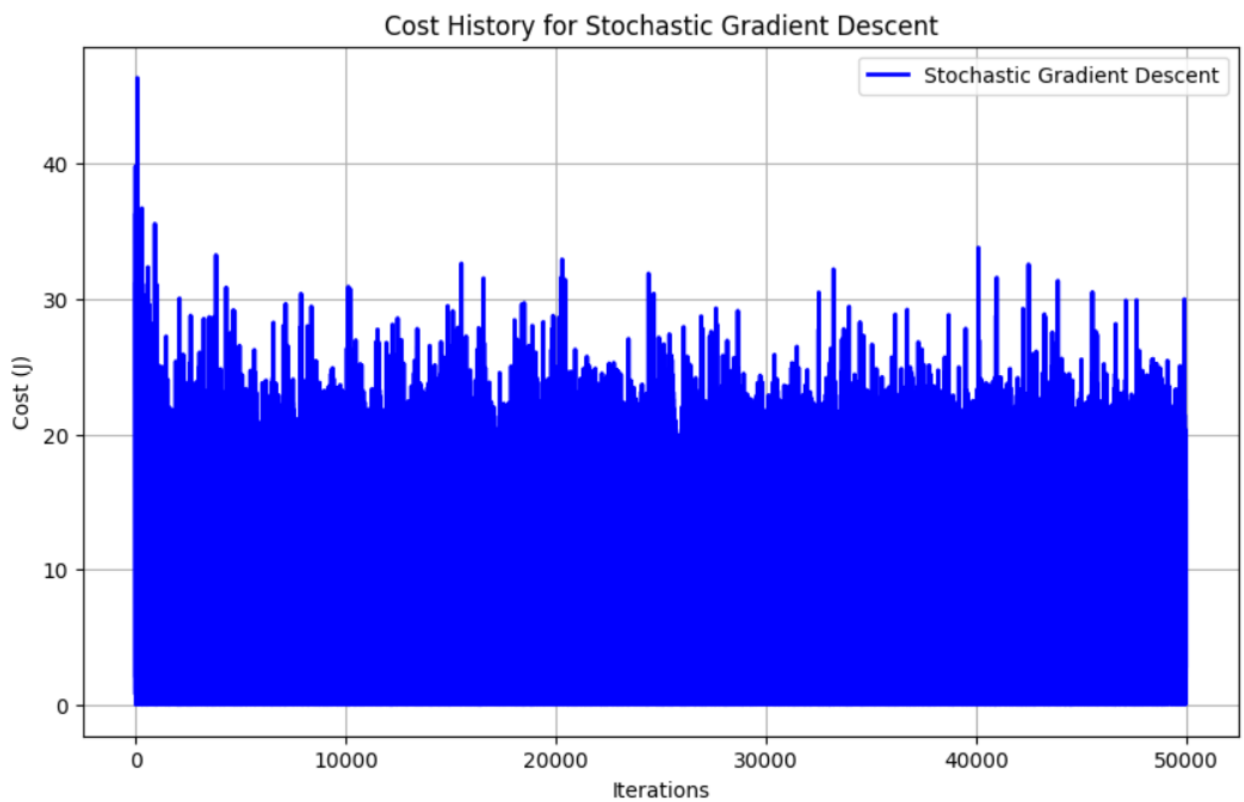
Practical Implications:

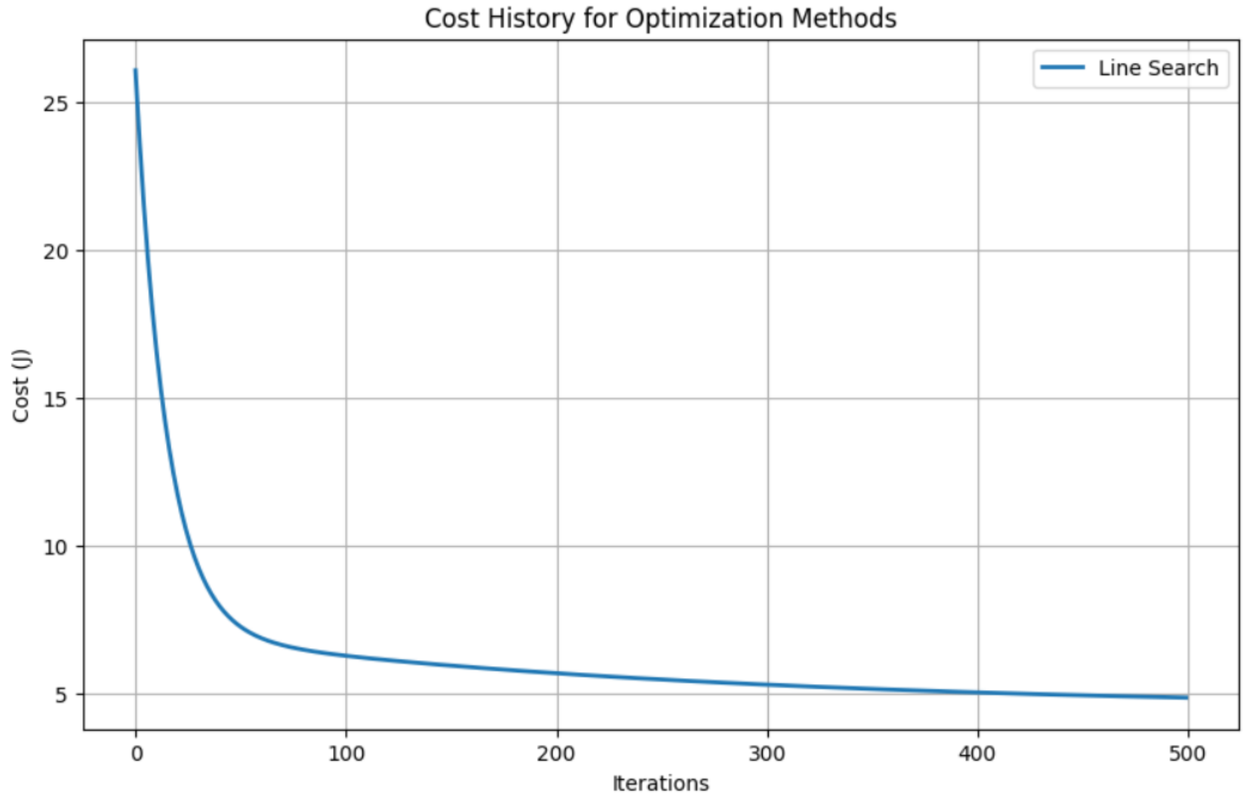
- **Line Search**: Useful when computational resources allow and the cost of determining an optimal step size is justified.
- **BGD**: Preferred when simplicity and lower computational overhead are critical.

Inference Summary

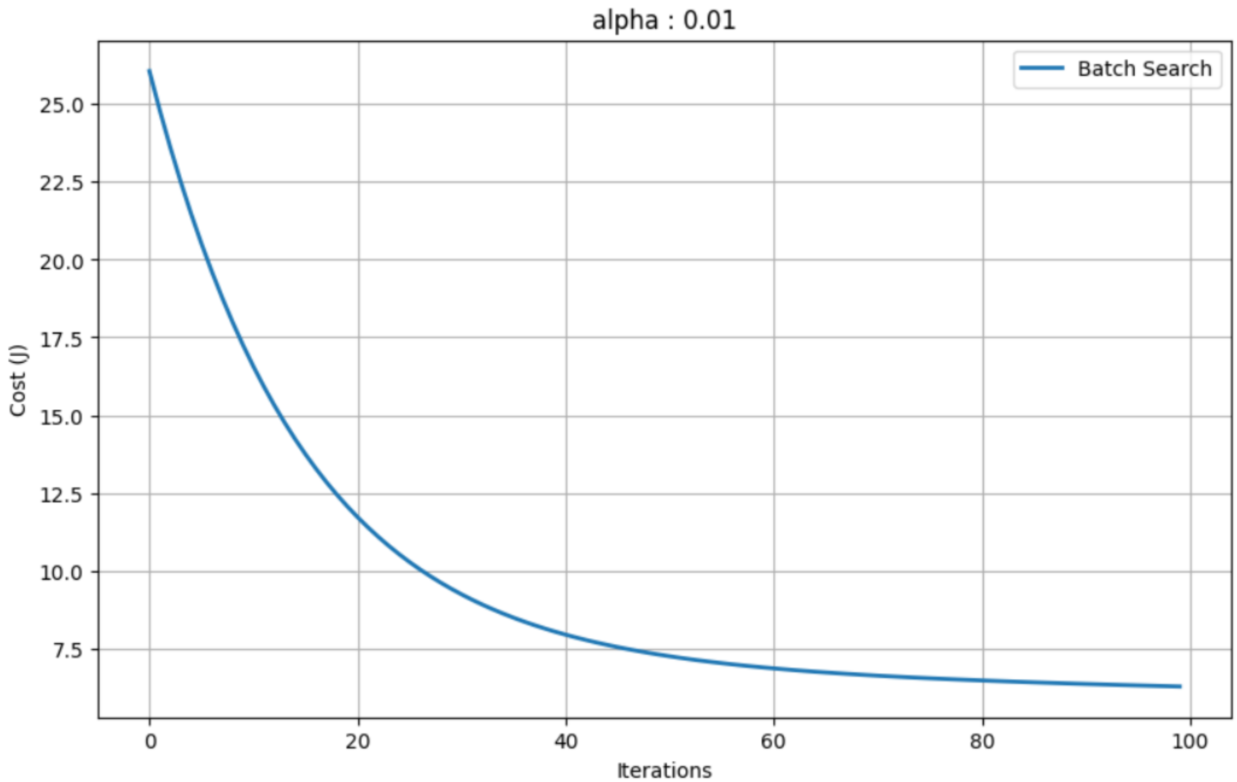
1. **SGD vs. BGD**: While SGD converges to a lower cost due to its ability to escape shallow local minima, its noisiness makes BGD more appealing for deterministic, stable convergence in smaller datasets.
2. **Impact of α**: A well-tuned α can significantly enhance convergence speed and stability. Adaptive techniques like line search mitigate the need for manual tuning, making them robust for practical applications.
3. **Line Search Advantage**: Line search achieves faster convergence by dynamically adapting the step size, but its computational cost makes it unsuitable for very large datasets.

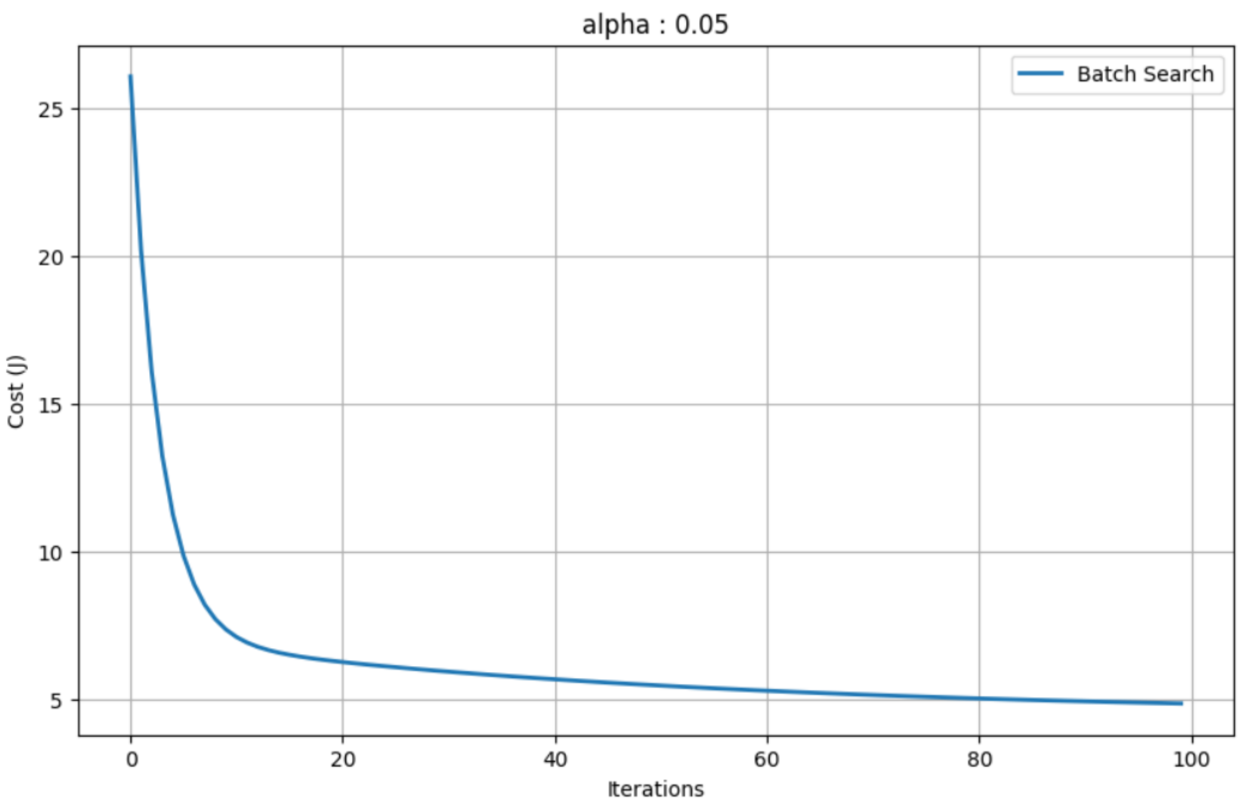

Cost History for Batch Gradient Descent
Batch Gradient Descent
Cost (J)
25
20
15
10
5
0
10000
20000
30000
40000
50000
Iterations

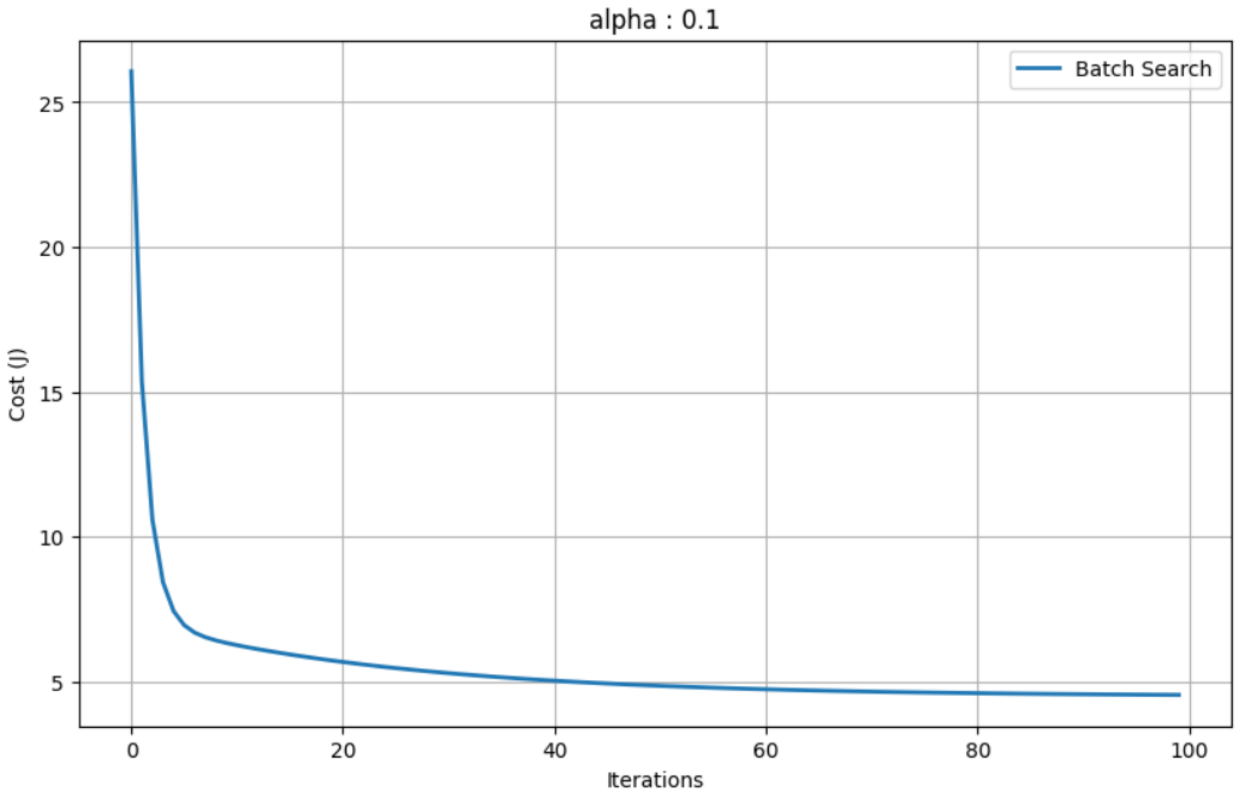

Cost History for Stochastic Gradient Descent
Stochastic Gradient Descent
Cost (J)
40
30
20
10
0
0
10000
20000
30000
40000
50000
Iterations

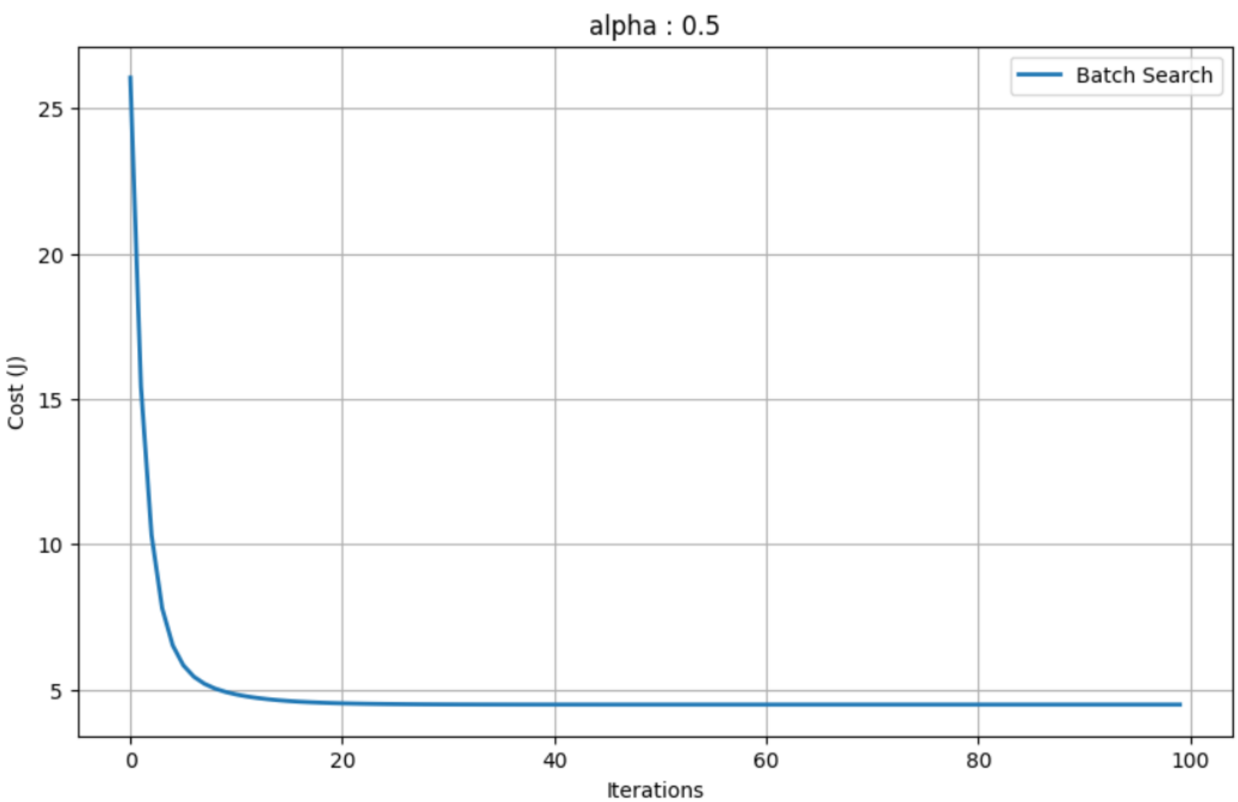


Cost History for Optimization Methods
Line Search
Cost (J)
25
20
15
10
5
0
100
200
300
400
500
Iterations


alpha : 0.01
Batch Search
Cost (J)
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
0
20
40
60
80
100
Iterations


alpha : 0.05
Batch Search
Cost (J)
25
20
15
10
5
0
20
40
60
80
100
Iterations


alpha : 0.1
Batch Search
Cost (J)
Iterations
0
20
40
60
80
100
5
10
15
20
25


alpha : 0.5
Batch Search
Cost (J)
Iterations
0
20
40
60
80
100
5
10
15
20
25