MDS6106 – Optimization and Modeling
Exercise 4

Zijin CAI

2024-11-22

Name: *Zijin CAI*
Student ID: *224040002*
Date: *2024-11-22*

Assignment A4.1: Implementing the Gradient Method

The implementation of `gradient_method()` return the final iterate x^k satisfying the stopping criterion, and consider the following input functions and parameters:

- `obj`, `grad` – function handles that calculate and return $f(x)$ and $\nabla f(x)$ at an input vector $x \in \mathbb{R}^n$
- `x0` – the initial point
- `tol` – the tolerance for the stopping criterion whenever $\|\nabla f(x^k)\| < \text{tol}$
- $\sigma, \gamma \in (0, 1)$ – the parameters for the backtracking and Armijo condition
- `alpha` – a function that returns a pre-defined, diminishing step size $\alpha_k \to 0$ ($\sum \alpha_k = \infty$)

The objective function:

$$\min_{x\in\mathbb{R}^2} f(x) := \frac{1}{2}x_1^4 - x_1^3 - x_1^2 + x_1^2x_2^2 + \frac{1}{2}x_2^4 - x_2^2 \tag{1}$$

The set of initial points:

$$\mathcal{X}^0 := \left\{ \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}, \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} -0.25 \\ -0.5 \end{bmatrix}, \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \right\} \tag{2}$$

the results and comparison to different methods:

The behavior and performance of the gradient method with backtracking line search, diminishing step size, and exact line search are shown in Figure 1, 2, and 3, respectively. The convergence points and the required number of iterations are print in the following verbatim block:

```
    Gradient Decent Method with Backtracking Line Search:
initial point: [-0.5  1. ], iterations: 14, convergence point: [ 2.00000044e+00 -7.15983494e-07]
initial point: [-0.5  0.5], iterations: 327, convergence point: [-0.00177274  0.99999842]
initial point: [-0.25 -0.5 ], iterations: 470, convergence point: [-0.00128325 -0.99999918]
initial point: [ 0.5 -0.5], iterations: 13, convergence point: [2.00000044e+00 7.15983494e-07]
initial point: [0.5 1. ], iterations: 11, convergence point: [ 1.99999975e+00 -5.21200527e-07]

    Gradient Method with Diminishing Step Size:
initial point: [-0.5  1. ], iterations: 48, convergence point: [2.00000094e+00 1.20629559e-36]
initial point: [-0.5  0.5], iterations: 8524, convergence point: [-0.00182567  0.99999833]
initial point: [-0.25 -0.5 ], iterations: 8502, convergence point: [-0.00182567 -0.99999833]
initial point: [ 0.5 -0.5], iterations: 48, convergence point: [ 1.99999906e+00 -1.36012921e-35]
initial point: [0.5 1. ], iterations: 48, convergence point: [2.00000094e+00 1.07763653e-37]

    Gradient Method with Exact Line Search:
initial point: [-0.5  1. ], iterations: 99, convergence point: [-0.00493936  0.99999811]
initial point: [-0.5  0.5], iterations: 99, convergence point: [-0.0048935   0.99995636]
initial point: [-0.25 -0.5 ], iterations: 99, convergence point: [-0.00465383 -0.99997491]
initial point: [ 0.5 -0.5], iterations: 11, convergence point: [2.00000001e+00 5.45313432e-08]
initial point: [0.5 1. ], iterations: 8, convergence point: [2.00000002e+00 4.26656946e-08]
```

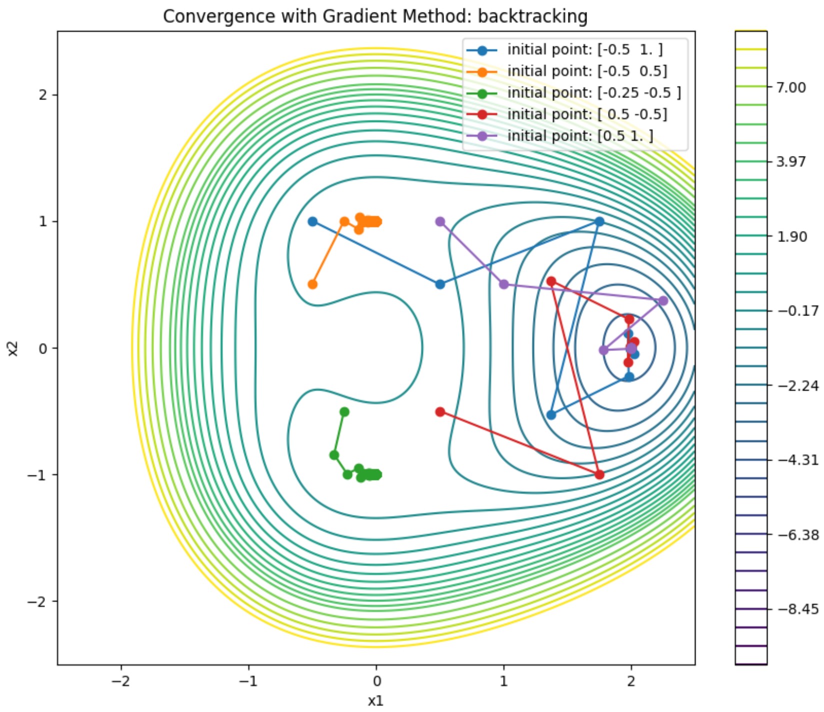

Figure 1: Gradient Method with Backtracking Line Search

- *Backtracking Line Search*: $(\sigma, \gamma) = (0.5, 0.1)$
  - show a significant variation in the number of iterations required for convergence (from 11 to 470)
  - generally converged to points very close to the optimal, with a high degree of precision
  - generally require fewer iterations to converge compared to the other two methods
  - especially few iterations for initial points like [-0.5, 1.0] and [0.5, -0.5]
  - demonstrate stability across different initial points and good convergence properties
- *Diminishing Step Size*: $\alpha_k = \frac{1}{\sqrt{k+2}}$
  - show good convergence but with more iterations required
  - require a large number of iterations for some initial points (i.e., 8502 for [-0.25, -0.5], and 8524 for [-0.5, 0.5]) but was able to converge to a point close to the optimal value with high precision
  - the convergence points were precise, with very small deviations from the expected values
- *Exact Line Search*: $maxit = 100, tol = 10^{-6}, a = 1$
  - show the most consistent convergence to points close to the expected optimal values with high precision
  - despite requiring more iterations for some points, provided the most precise convergence points
  - indicate a high level of efficiency of accuracy, especially for the initial points [0.5, -0.5] and [0.5, 1.0]

To conclude, each method has its strengths that the Backtracking Line Search is efficient in terms of iteration count and provides good precision, the Diminishing Step Size is stable but requires more iterations, especially for certain initial points, and the Exact Line Search offers the highest precision at the cost of more iterations, which may be acceptable where accuracy is paramount. The choice of method would depend on the specific requirements of the optimization problem, such as the balance between speed and precision.

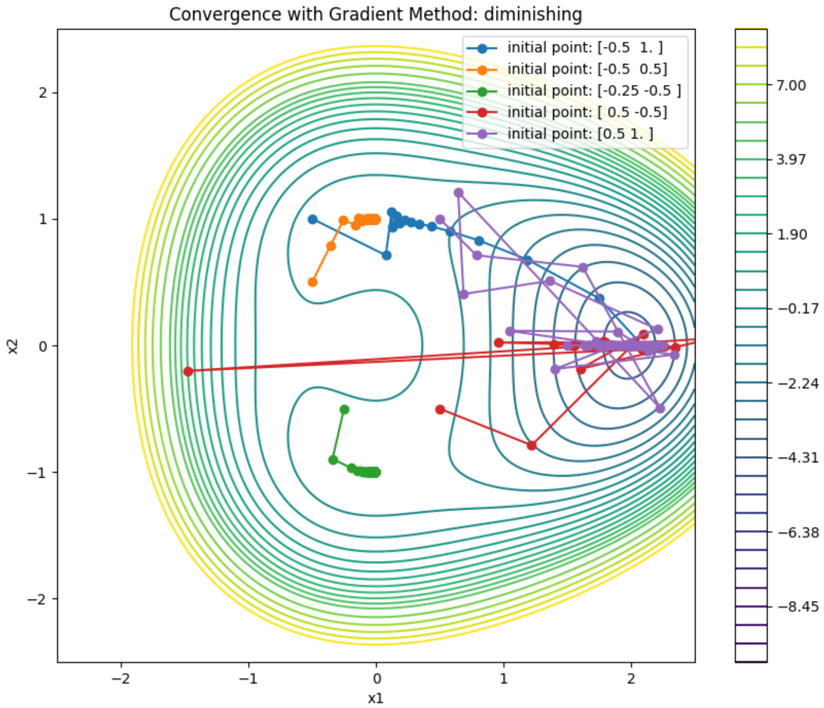

Figure 2: Gradient Method with Diminishing Step Size

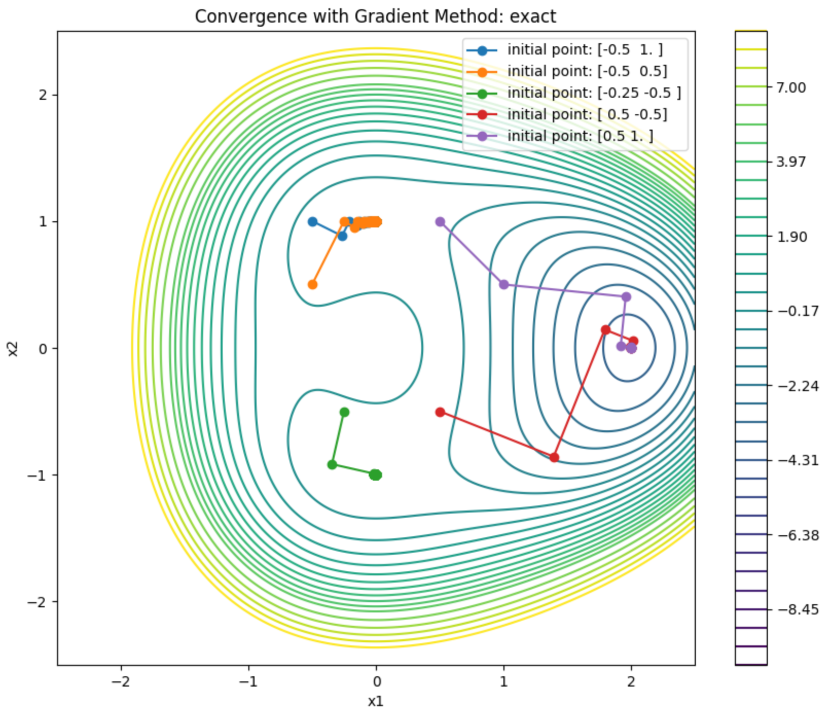

Figure 3: Gradient Method with Exact Line Search

Assignment A4.2: Inertial Gradient Method

The implementation of `inertial_gradient_method()` return the final iterate x^k satisfying the stopping criterion, and consider the following input functions and parameters:

- `obj`, `grad` – function handles that calculate and return $f(x)$ and $\nabla f(x)$ at an input vector $x \in \mathbb{R}^n$
- `x0` – the initial point
- `tol` – the tolerance for the stopping criterion whenever $\|\nabla f(x^k)\| < \text{tol}$
- $\beta \in (0, 1)$, $l > 0$ – the momentum parameter and an estimation of the Lipschitz constant

the results and comparison to different choice of β:

The behavior and performance of the inertial gradient method with different values of $\beta \in \{0.3, 0.5, 0.7, 0.9\}$ are shown in Figure 4, 5, 6 and 7, respectively. The convergence points and the required number of iterations are print in the following verbatim block:

```
    Inertial Gradient Method with beta 0.3:
initial point: [-0.5  1. ], iterations: 26, convergence point: [2.00000056e+00 2.23842110e-07]
initial point: [-0.5  0.5], iterations: 33, convergence point: [ 1.99999937e+00 -1.00254177e-06]
initial point: [-0.25 -0.5 ], iterations: 417, convergence point: [-0.00182343 -0.99999833]
initial point: [ 0.5 -0.5], iterations: 24, convergence point: [1.99999956e+00 1.15205964e-06]
initial point: [0.5 1. ], iterations: 24, convergence point: [1.99999921e+00 3.24514407e-07]
    Inertial Gradient Method with beta 0.5:
initial point: [-0.5  1. ], iterations: 39, convergence point: [ 1.99999921e+00 -9.27867489e-07]
initial point: [-0.5  0.5], iterations: 56, convergence point: [ 1.99999990e+00 -1.45093009e-07]
initial point: [-0.25 -0.5 ], iterations: 88, convergence point: [1.99999928e+00 1.05324181e-06]
initial point: [ 0.5 -0.5], iterations: 37, convergence point: [ 2.00000078e+00 -1.24151987e-07]
initial point: [0.5 1. ], iterations: 39, convergence point: [ 1.99999950e+00 -5.60399012e-07]
    Inertial Gradient Method with beta 0.7:
initial point: [-0.5  1. ], iterations: 105, convergence point: [2.00000031e+00 1.15222058e-06]
initial point: [-0.5  0.5], iterations: 88, convergence point: [2.00000019e+00 1.27561561e-06]
initial point: [-0.25 -0.5 ], iterations: 102, convergence point: [ 2.00000066e+00 -9.96531493e-07]
initial point: [ 0.5 -0.5], iterations: 77, convergence point: [ 2.00000021e+00 -1.17503102e-07]
initial point: [0.5 1. ], iterations: 79, convergence point: [ 2.00000005e+00 -3.50981713e-07]
    Inertial Gradient Method with beta 0.9:
initial point: [-0.5  1. ], iterations: 268, convergence point: [1.99999950e+00 3.90190143e-07]
initial point: [-0.5  0.5], iterations: 258, convergence point: [1.99999980e+00 5.66803197e-07]
initial point: [-0.25 -0.5 ], iterations: 276, convergence point: [2.00000055e+00 2.45681217e-07]
initial point: [ 0.5 -0.5], iterations: 297, convergence point: [ 1.99999918e+00 -4.92368826e-07]
initial point: [0.5 1. ], iterations: 5151, convergence point: [-0.00182565  0.99999833]
```

- $\beta = 0.3$ – This choice of β converges quickly for some initial points, requiring few as 25 iterations, while up to 418 iterations. The convergence points are generally close to the optimal, indicating effective optimization.
- $\beta = 0.5$ – This setting shows a balance between speed and stability, with iterations ranging from 38 to 89. The convergence points are very close to the optimal, suggesting good performance.
- $\beta = 0.7$ – This method converges in a moderate number of iterations (78 to 106), and the convergence points are very close to the optimal, indicating effective optimization.
- $\beta = 0.9$ – This high momentum setting shows a wide range of iterations (259 to 5152), with some cases converging very quickly and others taking many iterations. This suggests that while high momentum can speed up convergence, it may also lead to instability or require more iterations in some cases.

In conclusion, the Inertial Gradient Method with an appropriately chosen β value can be an effective optimization tool, offering a good balance between speed and stability. The choice of β should be determined by the specific characteristics of the optimization problem in practice.

comparison with standard gradient descent method:

- The standard gradient descent methods (Backtracking, Diminishing Step Size, Exact Line Search) were previously discussed and showed varying iteration counts and convergence points.
- The inertial gradient method with $\beta = 0.3$ and 0.5 generally required fewer iterations for convergence compared to $\beta = 0.7$ and 0.9, which suggests a lower momentum might be more effective in these cases.
- The convergence points for the inertial gradient method are generally close to the optimal, with some variations depending on the initial point and the value of β.
- Effect of $Beta$:
  - Lower Beta (0.3, 0.5) seems to provide a good balance between convergence speed and stability. The method converges relatively quickly, and the convergence points are close to the optimal.
  - Higher Beta (0.7, 0.9) can lead to oscillations or divergence, especially for $\beta = 0.9$ with the initial point [0.5, 1.0], taking 5152 iterations to converge, indicating potential instability with high momentum.

To conclude, The inertial gradient method can be effective, but the choice of β is crucial. Lower values of β (e.g., 0.3, 0.5) tend to provide faster convergence and stability. Higher β values (e.g., 0.7, 0.9) may lead to slower convergence or even divergence, depending on the initial point. The method's performance is competitive with standard gradient descent methods, offering a potential advantage in convergence speed and stability with appropriate momentum settings.

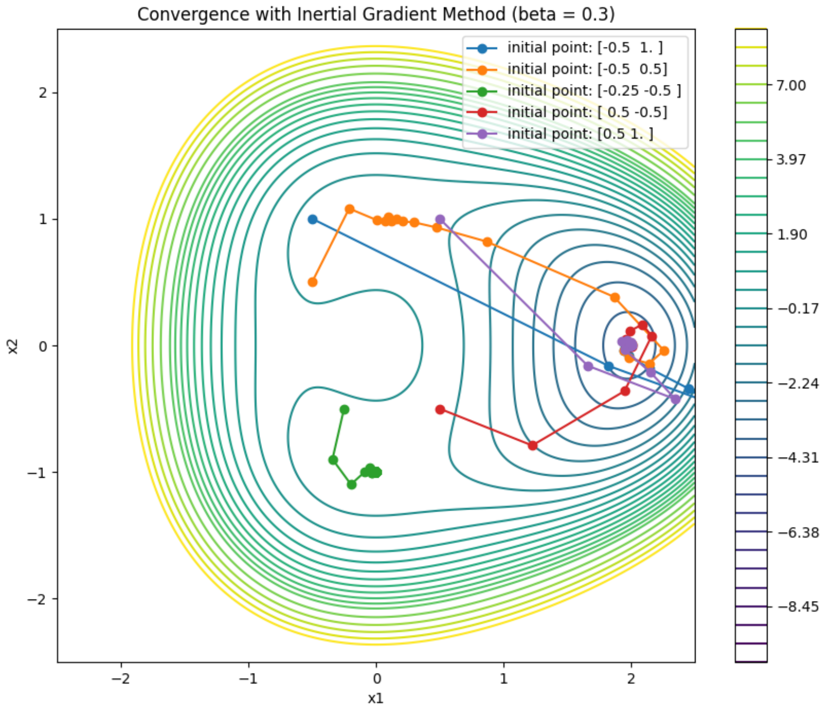

Figure 4: Inertial Gradient Method with $\beta = 0.3$

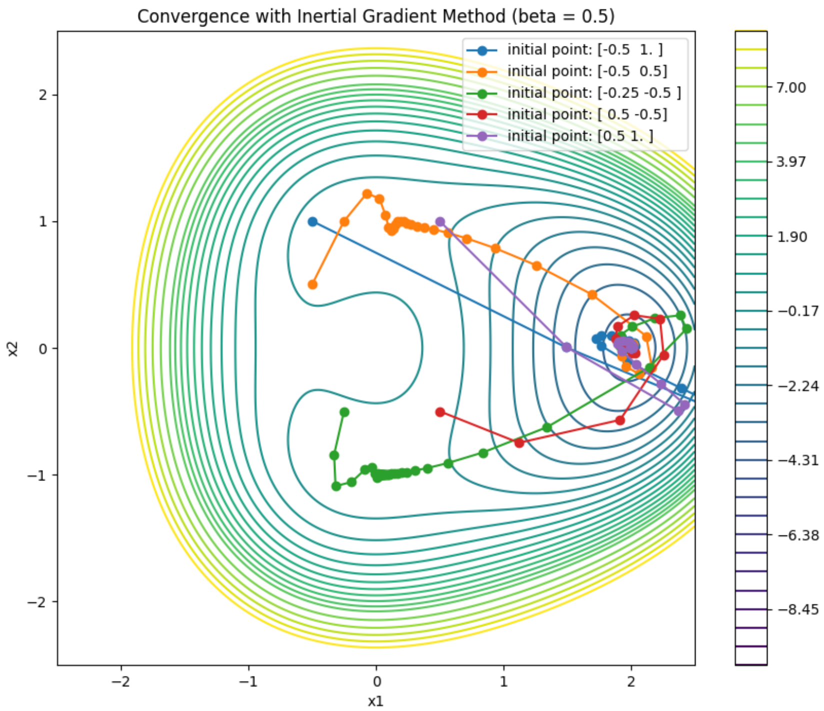

Figure 5: Inertial Gradient Method with $\beta = 0.5$

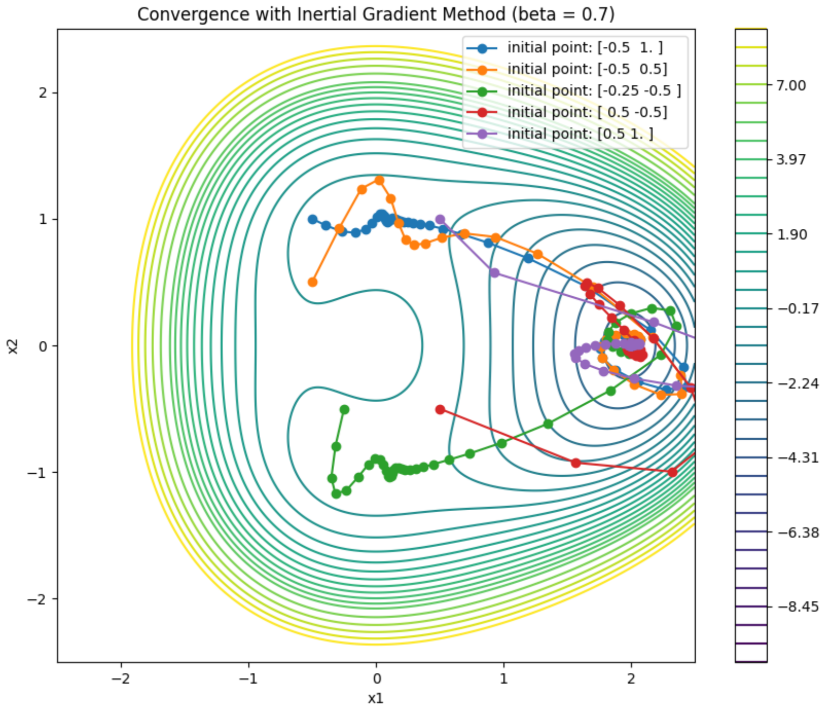

Figure 6: Inertial Gradient Method with $\beta = 0.7$

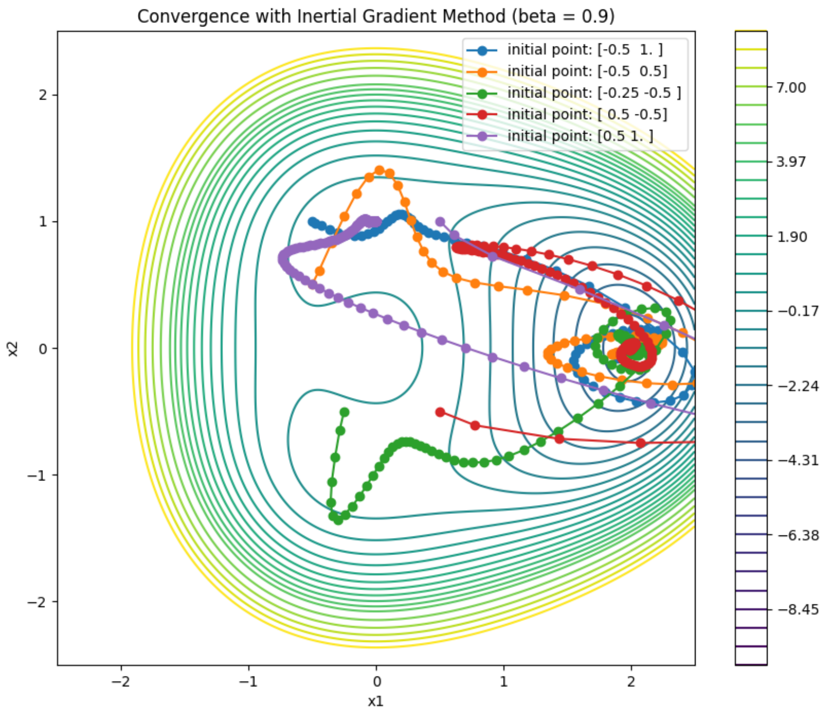

Figure 7: Inertial Gradient Method with $\beta = 0.9$