Optimization Methods for the Rosenbrock Function

Connor Leipelt

December 10, 2024

Introduction

- We were tasked with investigating the Rosenbrock function by using a multitude of optimization methods to find the global minimum of the fuction.
- The methods tested in this presentation are: Gradient Descent, Gradient Descent with Momentum, Nesterov Acceleration, AdaGrad, Adam, Newton's Method, and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method.

The Rosenbrock Function

- The Rosenbrock function is defined as:

$$f(x, y) = (a - x)^2 + b(y - x^2)^2$$

- For this project we will be taking, $a = 1$ and $b = 100$.
- It has a global minimum at $(x, y) = (a, a^2) = (1,1)$ for our case.
- (Note that for all methods I am using a 10^{-8} tolerance level.)

Gradient Descent Method

- Gradient Descent is an iterative optimization algorithm used to minimize a function.
- The update rule for Gradient Descent is:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)$$

  where α is the learning rate.
- Steps of the Gradient Descent method:
  1. Initialize the starting point $\mathbf{x}_0$.
  2. Compute the gradient $\nabla f(\mathbf{x}_k)$.
  3. Update the parameters using the update rule.
  4. Repeat until convergence.

Gradient Descent Results

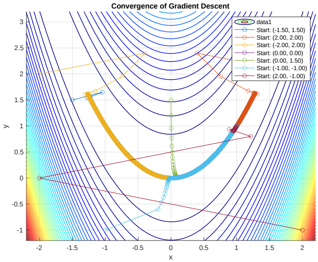


Figure: Graph of Gradient Descent

```
Command Window
>> Gradient_Descent
Starting point: (-1.50, 1.50)
Converged to: (0.999989, 0.999978)
Number of iterations: 26621
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (1.000011, 1.000022)
Number of iterations: 26508
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999989, 0.999978)
Number of iterations: 26667
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999989, 0.999978)
Number of iterations: 25540
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999989, 0.999978)
Number of iterations: 25516
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999989, 0.999978)
Number of iterations: 25571
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (0.999989, 0.999978)
Number of iterations: 20875
Function value at minimum: 0.000000

>>
```

Figure: Numerical results

Gradient Descent with Momentum

- Gradient Descent with Momentum accelerates the convergence by adding a fraction of the previous update to the current update.
- The update rule is:

$$\mathbf{v}_{k+1} = \gamma \mathbf{v}_k + \alpha \nabla f(\mathbf{x}_k)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{v}_{k+1}$$

  where γ is the momentum term and α is the learning rate.
- Steps of the Gradient Descent with Momentum method:
  1. Initialize the starting point $\mathbf{x}_0$ and $\mathbf{v}_0 = 0$.
  2. Compute the gradient $\nabla f(\mathbf{x}_k)$.
  3. Update the velocity and parameters using the update rules.
  4. Repeat until convergence.
- Note that this method depends heavily on the parameter γ and the initial conditions. Therefore we have graphs ranging from $\gamma = 0.3$ to $\gamma = 0.9$. We see divergence for $\gamma = 0.8, 0.9$

Gradient Descent with Momentum Results

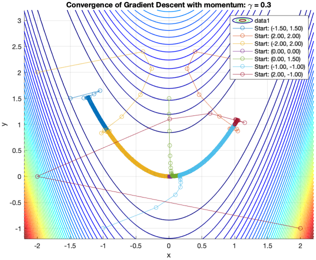


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.3$

```
>> Gradient_Descent_with_Momentum
Starting point: (-1.50, 1.50)
Converged to: (0.999218, 0.998434)
Number of iterations: 11178
Function value at minimum: 0.000001

Starting point: (2.00, 2.00)
Converged to: (0.999218, 0.998434)
Number of iterations: 7151
Function value at minimum: 0.000001

Starting point: (-2.00, 2.00)
Converged to: (0.999218, 0.998434)
Number of iterations: 10889
Function value at minimum: 0.000001

Starting point: (0.00, 0.00)
Converged to: (0.999218, 0.998434)
Number of iterations: 10433
Function value at minimum: 0.000001

Starting point: (0.00, 1.50)
Converged to: (0.999218, 0.998434)
Number of iterations: 10424
Function value at minimum: 0.000001

Starting point: (-1.00, -1.00)
Converged to: (0.999218, 0.998434)
Number of iterations: 10378
Function value at minimum: 0.000001

Starting point: (2.00, -1.00)
Converged to: (1.000783, 1.001570)
Number of iterations: 7014
Function value at minimum: 0.000001
```

Figure: Numerical results, $\gamma = 0.3$

Gradient Descent with Momentum Results

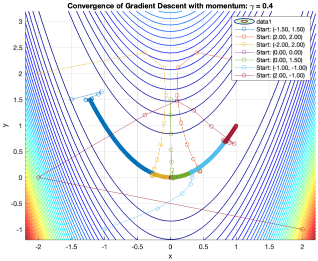


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.4$

```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999330, 0.998658)
Number of iterations: 9802
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (0.999330, 0.998658)
Number of iterations: 8988
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999330, 0.998658)
Number of iterations: 9254
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999330, 0.998658)
Number of iterations: 9171
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999330, 0.998658)
Number of iterations: 9167
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999330, 0.998658)
Number of iterations: 9040
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (0.999330, 0.998658)
Number of iterations: 7936
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.4$

Gradient Descent with Momentum Results

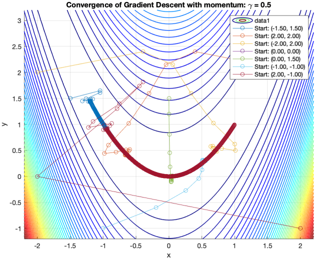
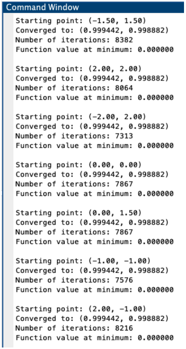


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.5$



```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999442, 0.998882)
Number of iterations: 8382
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (0.999442, 0.998882)
Number of iterations: 8064
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999442, 0.998882)
Number of iterations: 7313
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999442, 0.998882)
Number of iterations: 7867
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999442, 0.998882)
Number of iterations: 7867
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999442, 0.998882)
Number of iterations: 7576
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (0.999442, 0.998882)
Number of iterations: 8216
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.5$

Gradient Descent with Momentum Results

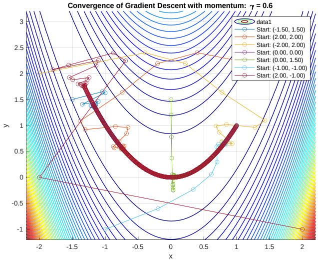


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.6$

```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999554, 0.999106)
Number of iterations: 6907
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (0.999554, 0.999106)
Number of iterations: 6707
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999554, 0.999106)
Number of iterations: 5800
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999554, 0.999106)
Number of iterations: 6511
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999554, 0.999106)
Number of iterations: 6514
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999554, 0.999106)
Number of iterations: 5749
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (0.999554, 0.999106)
Number of iterations: 7015
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.6$

Gradient Descent with Momentum Results

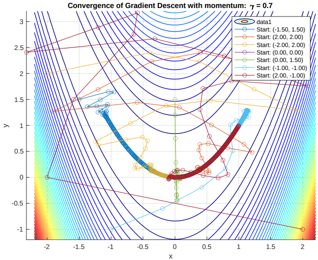
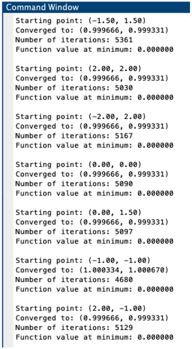


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.7$



```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999666, 0.999331)
Number of iterations: 5361
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (0.999666, 0.999331)
Number of iterations: 5030
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999666, 0.999331)
Number of iterations: 5167
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999666, 0.999331)
Number of iterations: 5090
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999666, 0.999331)
Number of iterations: 5097
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (1.000334, 1.000670)
Number of iterations: 4680
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (0.999666, 0.999331)
Number of iterations: 5129
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.7$

Gradient Descent with Momentum Results

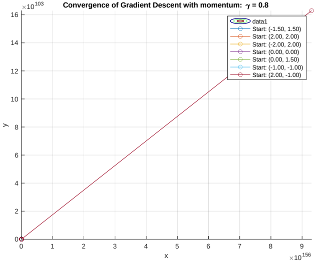


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.8$

```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999779, 0.999556)
Number of iterations: 3710
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (0.999778, 0.999556)
Number of iterations: 3512
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999778, 0.999556)
Number of iterations: 3544
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999779, 0.999556)
Number of iterations: 3576
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999778, 0.999556)
Number of iterations: 3591
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (1.000222, 1.000444)
Number of iterations: 4421
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (NaN, NaN)
Number of iterations: 30000
Function value at minimum: NaN
```

Figure: Numerical results, $\gamma = 0.8$

Gradient Descent with Momentum Results

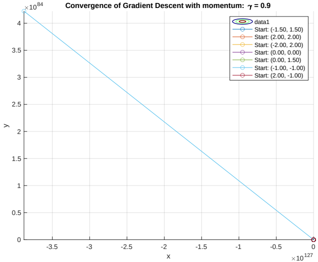


Figure: Graph of Gradient Descent with Momentum, $\gamma = 0.9$

```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999893, 0.999785)
Number of iterations: 1925
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (0.999893, 0.999785)
Number of iterations: 1950
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (0.999893, 0.999785)
Number of iterations: 1954
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999893, 0.999785)
Number of iterations: 1898
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999893, 0.999785)
Number of iterations: 1839
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (NaN, NaN)
Number of iterations: 30000
Function value at minimum: NaN

Starting point: (2.00, -1.00)
Converged to: (NaN, NaN)
Number of iterations: 30000
Function value at minimum: NaN

fx >>
```

Figure: Numerical results, $\gamma = 0.9$

Nesterov Acceleration

- Nesterov Acceleration is an optimization method that improves the convergence rate of gradient descent.
- The update rule is:

$$\mathbf{v}_{k+1} = \gamma \mathbf{v}_k + \alpha \nabla f(\mathbf{x}_k - \gamma \mathbf{v}_k)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{v}_{k+1}$$

  where γ is the momentum term and α is the learning rate.
- Steps of the Nesterov Acceleration method:
  1. Initialize the starting point $\mathbf{x}_0$ and $\mathbf{v}_0 = 0$.
  2. Compute the gradient $\nabla f(\mathbf{x}_k - \gamma \mathbf{v}_k)$.
  3. Update the velocity and parameters using the update rules.
  4. Repeat until convergence.
- Similarly to the Gradient Descent with Momentum method, this method depends heavily on the parameter γ and the initial conditions. Therefore we have graphs ranging from $\gamma = 0.3$ to $\gamma = 0.9$. We see divergence for $\gamma = 0.8, 0.9$

Nesterov Acceleration Results

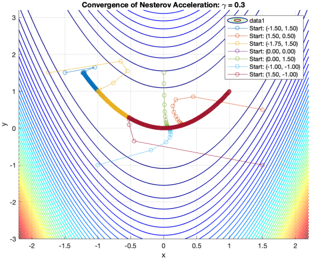


Figure: Graph of Nesterov Acceleration, $\gamma = 0.3$

```
Command Window
>> Nesterov_Acceleration
Starting point: (-1.50, 1.50)
Converged to: (0.999992, 0.999984)
Number of iterations: 19232
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999992, 0.999984)
Number of iterations: 18376
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999992, 0.999984)
Number of iterations: 19013
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999992, 0.999984)
Number of iterations: 18501
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999992, 0.999984)
Number of iterations: 18489
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999992, 0.999984)
Number of iterations: 18475
Function value at minimum: 0.000000

Starting point: (1.50, -1.00)
Converged to: (0.999992, 0.999984)
Number of iterations: 18705
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.3$

Nesterov Acceleration Results

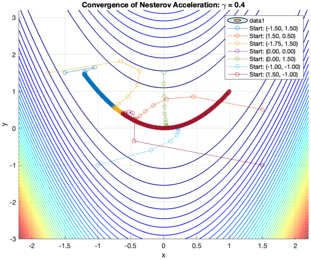


Figure: Graph of Nesterov Acceleration, $\gamma = 0.4$

```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999993, 0.999987)
Number of iterations: 16702
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999993, 0.999987)
Number of iterations: 16230
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999993, 0.999987)
Number of iterations: 16361
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999993, 0.999987)
Number of iterations: 16087
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999993, 0.999987)
Number of iterations: 16079
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999993, 0.999987)
Number of iterations: 16017
Function value at minimum: 0.000000

Starting point: (1.50, -1.00)
Converged to: (0.999993, 0.999987)
Number of iterations: 16305
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.4$

Nesterov Acceleration Results

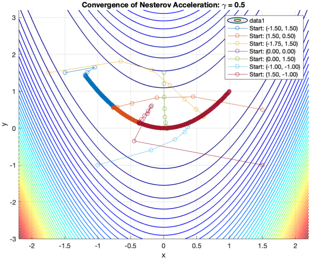
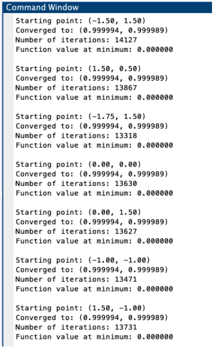


Figure: Graph of Nesterov Acceleration, $\gamma = 0.5$



Command Window

```
Starting point: (-1.50, 1.50)
Converged to: (0.999994, 0.999989)
Number of iterations: 14127
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999994, 0.999989)
Number of iterations: 13867
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999994, 0.999989)
Number of iterations: 13318
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999994, 0.999989)
Number of iterations: 13630
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999994, 0.999989)
Number of iterations: 13627
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999994, 0.999989)
Number of iterations: 13471
Function value at minimum: 0.000000

Starting point: (1.50, -1.00)
Converged to: (0.999994, 0.999989)
Number of iterations: 13731
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.5$

Nesterov Acceleration Results

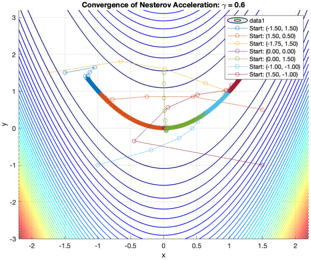


Figure: Graph of Nesterov Acceleration, $\gamma = 0.6$

```
Command Window
Starting point: (-1.50, 1.50)
Converged to: (0.999996, 0.999991)
Number of iterations: 11498
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999996, 0.999991)
Number of iterations: 11401
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999996, 0.999991)
Number of iterations: 9503
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999996, 0.999991)
Number of iterations: 11121
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999996, 0.999991)
Number of iterations: 11121
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999996, 0.999991)
Number of iterations: 10737
Function value at minimum: 0.000000

Starting point: (1.50, -1.00)
Converged to: (1.000004, 1.000009)
Number of iterations: 10810
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.6$

Nesterov Acceleration Results

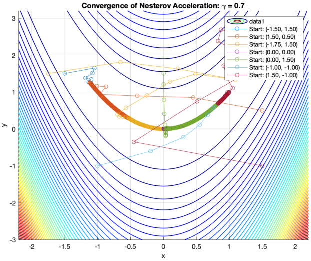


Figure: Graph of Nesterov Acceleration, $\gamma = 0.7$

Command Window

```
Starting point: (-1.50, 1.50)
Converged to: (0.999997, 0.999993)
Number of iterations: 8798
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999997, 0.999993)
Number of iterations: 8824
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999997, 0.999993)
Number of iterations: 8636
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999997, 0.999993)
Number of iterations: 8544
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999997, 0.999993)
Number of iterations: 8548
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999997, 0.999993)
Number of iterations: 5168
Function value at minimum: 0.000000

Starting point: (1.50, -1.00)
Converged to: (0.999997, 0.999993)
Number of iterations: 7958
Function value at minimum: 0.000000
```

Figure: Numerical results, $\gamma = 0.7$

Nesterov Acceleration Results

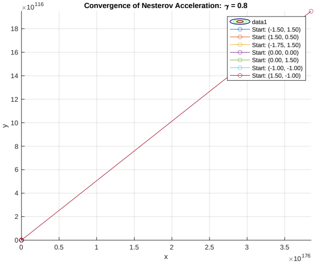


Figure: Graph of Nesterov Acceleration, $\gamma = 0.8$

Command Window

```
Starting point: (-1.50, 1.50)
Converged to: (0.999998, 0.999996)
Number of iterations: 6001
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999998, 0.999996)
Number of iterations: 6035
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999998, 0.999996)
Number of iterations: 5903
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999998, 0.999996)
Number of iterations: 5871
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999998, 0.999996)
Number of iterations: 5878
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999998, 0.999996)
Number of iterations: 5784
Function value at minimum: 0.000000

Starting point: (1.50, -1.00)
Converged to: (NaN, NaN)
Number of iterations: 30000
Function value at minimum: NaN
```

Figure: Numerical results, $\gamma = 0.8$

Nesterov Acceleration Results

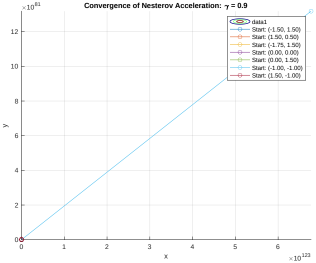


Figure: Graph of Nesterov Acceleration, $\gamma = 0.9$

Command Window

```
Starting point: (-1.50, 1.50)
Converged to: (0.999999, 0.999998)
Number of iterations: 3030
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (0.999999, 0.999998)
Number of iterations: 3103
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999999, 0.999998)
Number of iterations: 3027
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999999, 0.999998)
Number of iterations: 3018
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999999, 0.999998)
Number of iterations: 3030
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (NaN, NaN)
Number of iterations: 30000
Function value at minimum: NaN

Starting point: (1.50, -1.00)
Converged to: (NaN, NaN)
Number of iterations: 30000
Function value at minimum: NaN

fx >>
```

Figure: Numerical results, $\gamma = 0.9$

AdaGrad Method

- AdaGrad is an adaptive learning rate optimization algorithm.
- The update rule is:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{G_t + \epsilon}} \nabla f(\theta_t)$$

  where G_t is the sum of the squares of the gradients up to time t, and ϵ is a small constant to prevent division by zero.
- Steps of the AdaGrad method:
  1. Initialize the starting point θ_0 and $G_0 = 0$.
  2. Compute the gradient $\nabla f(\theta_t)$.
  3. Accumulate the squared gradients: $G_t = G_{t-1} + \nabla f(\theta_t)^2$.
  4. Update the parameters using the update rule.
  5. Repeat until convergence.
- This method performed very well in accuracy of the global minimum, but as you will see in the next slide, the rate of convergence can be extremely slow for specific initial conditions.

AdaGrad Results

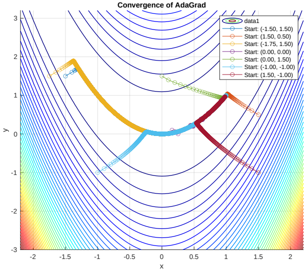
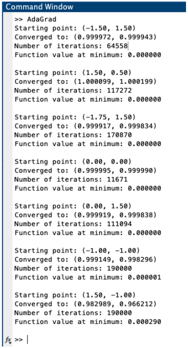


Figure: Graph of AdaGrad, all but two initial conditions converge in given time step.



Command Window

```
>> AdaGrad
Starting point: (-1.50, 1.50)
Converged to: (0.999972, 0.999943)
Number of iterations: 64558
Function value at minimum: 0.000000

Starting point: (1.50, 0.50)
Converged to: (1.000099, 1.000199)
Number of iterations: 117272
Function value at minimum: 0.000000

Starting point: (-1.75, 1.50)
Converged to: (0.999917, 0.999834)
Number of iterations: 170870
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (0.999995, 0.999990)
Number of iterations: 11671
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (0.999919, 0.999838)
Number of iterations: 111094
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (0.999149, 0.998296)
Number of iterations: 190000
Function value at minimum: 0.000001

Starting point: (1.50, -1.00)
Converged to: (0.982989, 0.966212)
Number of iterations: 190000
Function value at minimum: 0.000290

>>
```

Figure: Numerical results

AdaGrad Results

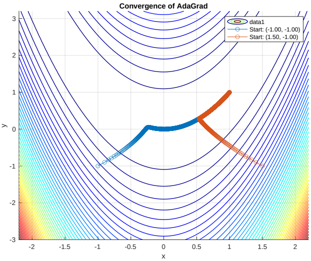


Figure: Graph of AdaGrad for the two longest convergence times ($246,673$ and $461,378$ iterations).

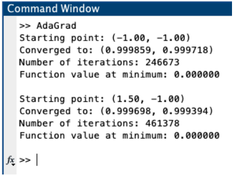


Figure: Numerical results

Adam Method

- Adam (Adaptive Moment Estimation) is an optimization algorithm that combines the advantages of AdaGrad and RMSProp.
- The update rules are:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla f(\theta_t)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla f(\theta_t))^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where α is the learning rate, β_1 and β_2 are the decay rates for the moment estimates, and ϵ is a small constant to prevent division by zero.

Adam Method continued

- Steps of the Adam method:
  1. Initialize the starting point θ_0, $m_0 = 0$, and $v_0 = 0$.
  2. Compute the gradient $\nabla f(\theta_t)$.
  3. Update the biased first moment estimate m_t.
  4. Update the biased second moment estimate v_t.
  5. Compute bias-corrected first moment estimate $\hat{m}_t$.
  6. Compute bias-corrected second moment estimate $\hat{v}_t$.
  7. Update the parameters using the update rule.
  8. Repeat until convergence.
- This method performed the best out of all of the optimization methods studied in this project. Newton's method is also very good, but requires knowledge about the Hessian which may be difficult.

Adam Results

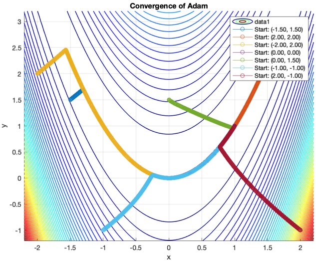


Figure: Graph of Adam

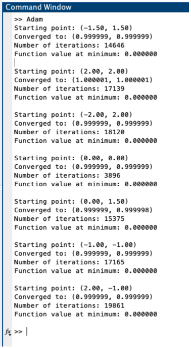


Figure: Numerical results

Newton's Method

- Newton's Method uses second-order derivatives to find the minimum.
- The update rule is:

$$\theta_{t+1} = \theta_t - \left(\nabla^2 f(\theta_t)\right)^{-1} \nabla f(\theta_t)$$

- Steps of Newton's Method:
  1. Initialize the starting point θ_0.
  2. Compute the gradient $\nabla f(\theta_t)$.
  3. Compute the Hessian matrix $\nabla^2 f(\theta_t)$.
  4. Update the parameters using the update rule:

$$\theta_{t+1} = \theta_t - \left(\nabla^2 f(\theta_t)\right)^{-1} \nabla f(\theta_t)$$

  5. Repeat until convergence.
- Note: The computation of the Hessian matrix can be computationally expensive for high-dimensional problems.

Newton's Method Results

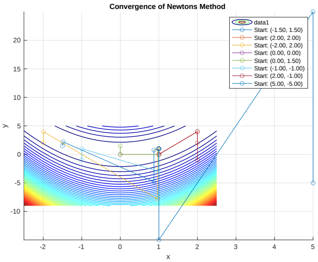


Figure: Graph of Newton's Method: Note that a farther point from the global minimum was chosen to show the ability of Newton's Method.

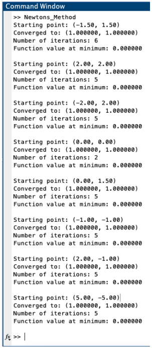


Figure: Numerical results

Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method

- The BFGS method is an iterative method for solving unconstrained nonlinear optimization problems.
- It is a quasi-Newton method, which means it approximates the Hessian matrix.
- The update rule is:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{H}_k^{-1} \nabla f(\mathbf{x}_k)$$

  where $\mathbf{H}_k$ is an approximation to the Hessian matrix.
- Steps of the BFGS method:
  1. Initialize the starting point $\mathbf{x}_0$ and $\mathbf{H}_0 = \mathbf{I}$ (identity matrix).
  2. Compute the gradient $\nabla f(\mathbf{x}_k)$.
  3. Compute the search direction: $\mathbf{p}_k = -\mathbf{H}_k \nabla f(\mathbf{x}_k)$.
  4. Perform a line search to find the step size α_k.
  5. Update the parameters: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$.
  6. Update the Hessian approximation:

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{\mathbf{H}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{H}_k}{\mathbf{s}_k^T \mathbf{H}_k \mathbf{s}_k}$$

     where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$.
  7. Repeat until convergence.

BFGS Method Results

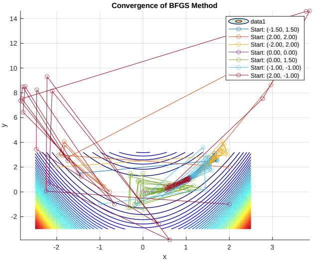


Figure: Graph of BFGS

```
Command Window
>> BFGS
Starting point: (-1.50, 1.50)
Converged to: (1.000000, 1.000000)
Number of iterations: 305
Function value at minimum: 0.000000

Starting point: (2.00, 2.00)
Converged to: (1.000000, 1.000000)
Number of iterations: 297
Function value at minimum: 0.000000

Starting point: (-2.00, 2.00)
Converged to: (1.000000, 1.000000)
Number of iterations: 313
Function value at minimum: 0.000000

Starting point: (0.00, 0.00)
Converged to: (1.000000, 1.000000)
Number of iterations: 270
Function value at minimum: 0.000000

Starting point: (0.00, 1.50)
Converged to: (1.000000, 1.000000)
Number of iterations: 327
Function value at minimum: 0.000000

Starting point: (-1.00, -1.00)
Converged to: (1.000000, 1.000000)
Number of iterations: 270
Function value at minimum: 0.000000

Starting point: (2.00, -1.00)
Converged to: (1.000000, 1.000000)
Number of iterations: 328
Function value at minimum: 0.000000

>>
```

Figure: Numerical results

Conclusion

- There does not exist one perfect method. Different optimization methods have their own strengths and weaknesses.
- The choice of method depends on the specific problem and computational resources.
- The Rosenbrock function is a useful benchmark for testing optimization algorithms and to see the steps of these methods.

Thank you!