IE 684 Lab 11: Optimization Algorithms
Gradient Descent, Momentum, and NAG

Instructions

- Solve all problems independently. Seek help from instructors/TAs if needed.
- Use Python (NumPy, Matplotlib) for implementations.
- Submit separate notebooks for each exercise:
  - `YOURROLLNUMBER_IE684_Lab05_Ex1.ipynb`
  - `YOURROLLNUMBER_IE684_Lab05_Ex2.ipynb`
- Include LaTeX equations for explanations, plots for results, and comments on observations.
- Zip all files as `YOURROLLNUMBER_IE684_Lab05.zip` and upload to Moodle.

Submission Guidelines

- Include LaTeX equations for all derivations.
- Label plots clearly (title, axes, legend).
- Comment on observations (e.g., "NAG converges faster due to anticipatory updates").
- Penalties for incorrect naming conventions.

Exercise 1: Mathematical Formulation (25 marks)

[**R**] Write the mathematical formulation for the following scenario to minimize the overall supply chain cost. There are two factories [P1, P2] which produce three products [PD1, PD2, PD3]. These products are transported from the factories to depots. There are five depots [D1, D2, D3, D4, D5] from where the products are delivered to ten customers [C1, C2, C3, C4, C5, C6, C7, C8, C9, C10]. While writing the formulation, the following parameters need to be considered :

- Production cost per product at factories.
- Transportion cost for factory to depot.
- Transportion cost for depot to customer.
- Depot capacity.
- Demand at customer location.

Exercise 2: Gradient Descent Variants (25 marks)

Consider the function:

$$f(\mathbf{x}) = 3x_1^2 + 2x_1x_2 + 3x_2^2 - 4x_1 - 4x_2$$

1. Theoretical Analysis (5 marks)

- Find the minimizer $\mathbf{x}^*$, minimum value $f(\mathbf{x}^*)$, and confirm if it's a global minimum.
- Compute the Hessian $\nabla^2 f(\mathbf{x})$. Is $f(\mathbf{x})$ convex? Justify.

2. Implementation (10 marks)

- Implement **Gradient Descent (GD)** with fixed step size $\eta = 0.1$. Use $\mathbf{x}_0 = (5, 5)$, $\tau = 10^{-6}$.
- Plot $\log(\|\mathbf{x}_k - \mathbf{x}^*\|_2)$ vs. iterations. Comment on convergence rate.
- Repeat for $\eta \in \{0.01, 0.05, 0.2\}$. Plot all trajectories on the same graph. Compare results.

3. Momentum and NAG (10 marks)

- Implement **Momentum GD** ($\beta = 0.9$) and **NAG** ($\beta = 0.9$) with $\eta = 0.1$.
- Plot $\log(|f(\mathbf{x}_k) - f(\mathbf{x}^*)|)$ for GD, Momentum, NAG. Compare convergence.
- Vary $\beta \in \{0.5, 0.9, 0.99\}$. Plot error vs. iterations for each β. Analyze the effect of momentum.

Algorithm 1 Gradient Descent with Momentum

Require: Initial point $\mathbf{x}_0$, step size η, momentum β, tolerance τ

Initialize $\mathbf{v}_0 = 0$, $k = 0$
while $\|\nabla f(\mathbf{x}_k)\|_2 > \tau$ **do**
    $\mathbf{v}_{k+1} = \beta \mathbf{v}_k - \eta \nabla f(\mathbf{x}_k)$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{v}_{k+1}$
    $k = k + 1$
end whilereturn $\mathbf{x}_k$

Algorithm 2 Nesterov Accelerated Gradient (NAG)

Require: Initial point $\mathbf{x}_0$, step size η, momentum β, tolerance τ

Initialize $\mathbf{v}_0 = 0$, $k = 0$
while $\|\nabla f(\mathbf{x}_k)\|_2 > \tau$ **do**
    $\mathbf{y}_k = \mathbf{x}_k + \beta \mathbf{v}_k$
    $\mathbf{v}_{k+1} = \beta \mathbf{v}_k - \eta \nabla f(\mathbf{y}_k)$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{v}_{k+1}$
    $k = k + 1$
end whilereturn $\mathbf{x}_k$

Exercise 3: Comparative Analysis (25 marks)

Consider the Rosenbrock function:

$$f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

1. GD vs. Momentum vs. NAG (15 marks)

- Implement GD ($\eta = 0.001$), Momentum GD ($\eta = 0.001, \beta = 0.9$), and NAG ($\eta = 0.001, \beta = 0.9$).
- Use $\mathbf{x}_0 = (-1.5, 2)$, $\tau = 10^{-6}$. Plot trajectories on contour plots.
- Record iterations to convergence. Tabulate results and explain differences.

2. Step Size Sensitivity (10 marks)

- For GD, Momentum, NAG, test $\eta \in \{0.0001, 0.001, 0.01, 0.1\}$.
- Plot final error vs. η. Identify "optimal" η for each method.
- Discuss trade-offs (speed vs. stability).

Exercise 4: Difference between the GD with Momentum and NAG

[**R**] Explain the difference between the Gradient Descent with Momentum and Nesterov Accelerated Gradient (NAG)

Reading Material

Click the link for finding the supplimentary material.