AI505/AI801, Optimization – Exercise Sheet 04

2026-03-04

Exercises with the symbol $^+$ are to be done at home before the class. Exercises with the symbol * will be tackled in class. The remaining exercises are left for self training after the exercise class. Some exercises are from the text book and the number is reported. They have the solution at the end of the book.

Exercise 1 * (4.1)

Find examples where each of the four termination conditions would not work individually, showing the importance of having more than one.

Exercise 2 * (6.1)

What advantage does second-order information provide about the point of convergence that first-order information lacks?

Exercise 3 * (6.2)

When would we use Newton's method instead of the bisection method for the task of finding roots in one dimension?

Exercise 4 * (6.4, 6.9)

Apply Newton's method to $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T H \boldsymbol{x}$ starting from $\boldsymbol{x}_0 = [1, 1]$. What have you observed? Use H as follows:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1000 \end{bmatrix}$$

Next, apply gradient descent to the same optimization problem by stepping with the unnormalized gradient. Do two steps of the algorithm. What have you observed? Finally, apply the conjugate gradient method. How many steps do you need to converge?

Repeat the exercise for:

$$f(\boldsymbol{x}) = (x_1 + 1)^2 + (x_2 + 3)^2 + 4.$$

starting at the origin.

Note that $H = A$, hence we could have derived A also by calculating the Hessian.

Exercise 5 * (6.5)

Compare Newton's method and the secant method on $f(x) = x^2 + x^4$, with $x_1 = -3$ and $x_0 = -4$. Run each method for 10 iterations. Make two plots:

1. Plot f vs. the iteration for each method.

2. Plot f' vs. x. Overlay the progression of each method, drawing lines from $(x_i, f'(x_i))$ to $(x_{i+1}, 0)$ to $(x_{i+1}, f'(x_{i+1}))$ for each transition.

What can we conclude about this comparison?

Exercise 6 Strongly Convex Functions

Consider:

$$f(x) = x^4.$$

- Show that f is convex.

- Compute $f''(x)$.

- Show that f is not strongly convex on $\mathbb{R}$.

- Explain geometrically why the function fails to be strongly convex.

Exercise 7 Strongly Convex Functions

Recall that for a symmetric matrix A,
$$A \succeq 0$$
means that A is positive semidefinite, i.e.

$$\boldsymbol{x}^\top A \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Hence, for Q a symmetric matrix, and I the identity matrix, the condition

$$Q \succeq cI$$

means:

$$Q - cI \succeq 0.$$

In scalar form this means:

$$\boldsymbol{x}^T (Q - cI) \boldsymbol{x} = \boldsymbol{x}^\top Q \boldsymbol{x} - c\|\boldsymbol{x}\|^2 \geq 0.$$

So, in its easiest form, the condition is:

$$\boldsymbol{x}^\top Q \boldsymbol{x} \geq c\|\boldsymbol{x}\|^2 \quad \forall \boldsymbol{x}.$$

Assume f is a twice-differentiable function (ie, $f \in C^2(\mathbb{R}^n)$). Show that:

$$f \text{ is } c\text{-strongly convex} \iff \nabla^2 f(\boldsymbol{x}) \succeq cI \text{ for all } \boldsymbol{x} \in \mathbb{R}^n.$$

(Hint: Use Taylor's theorem with integral remainder.)

Exercise 8

Implement gradient descent for:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x}$$

where

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1/\gamma \end{bmatrix}.$$

- Run experiments for $\gamma = 1, 10, 100, 1000$.
- Plot convergence.
- Relate the behavior to the condition number.
- Observe how increasing strong convexity improves convergence.

Exercise 9

Show that strong convexity is equivalent to strong monotonicity of the gradient:

$$(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}))^\top (\boldsymbol{x} - \boldsymbol{y}) \geq c\|\boldsymbol{x} - \boldsymbol{y}\|^2 \quad \forall \boldsymbol{x}, \boldsymbol{y}.$$