ECE 509 (Spring'25): Homework #6

80 points

Problem 1 (10 points): Complete Exercise 9.6 from Boyd and Vandenberghe.

Problem 2 (10 points): Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable, strongly convex function. Suppose the Hessian of f at each point is of the form $\nabla^2 f(x) = D + uv^\top$, where $D \in \mathbb{R}^{n\times n}$ is a diagonal matrix and $u, v \in \mathbb{R}^n$.

Recall the matrix inversion lemma (also known as the Sherman–Morrison–Woodbury formula), which states that for an invertible matrix $A \in \mathbb{R}^{n\times n}$, and vectors $u, v \in \mathbb{R}^n$,

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}.$$

What is the best possible cost (in big-O notation) of each iteration of Newton's method for minimizing f, expressed as a function of n? You may assume that computing the gradient takes $O(n)$ time.

Problem 3 (10 points): Exercise 9.12 from *Additional Exercises for Convex Optimization* by Boyd and Vandenberghe. Justify your answers fully. You will need to use the solution to the previous problem to support your reasoning. In particular, compute the Hessian explicitly, even though the problem statement does not ask for it.

Problem 4 (50 points): Consider the function

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}.$$

You will implement and compare various descent methods starting from the same initial point, using backtracking line search with parameters $\alpha = 0.1$ and $\beta = 0.7$. The methods to be compared are:

- Steepest descent in the ℓ_1 norm, using the direction

  $$\Delta x_{\text{sd}} = -\frac{\partial f(x)}{\partial x_i} e_i,$$

  where i is an index for which $|\nabla f(x)|_\infty = |[\nabla f(x)]_i|$ and e_i is the ith standard basis vector.
- Steepest descent in the ℓ_∞ norm, using the direction

  $$\Delta x_{\text{sd}} = -\|\nabla f(x)\|_1 \operatorname{sign}(\nabla f(x)),$$

  where the sign is taken componentwise.
- Steepest descent in a quadratic P-norm, where

  $$P_1 = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}.$$

- Gradient descent using the standard Euclidean norm.
- Newton's method, using the exact Hessian of the function.

For each method, overlay the iterates on the contour plot of $f(x_1, x_2)$, and also produce a plot showing the function value gap $f(x^{(k)}) - p^\star$ versus the iteration number on a semi-log scale (log scale on the y-axis and linear on the x-axis).

Comment on which steepest descent variant performs best and which performs worst. Justify your answer based on the observed behavior and geometry of the level sets. Compare the performance of these norm-based steepest descent methods with classical gradient descent and Newton's method. Explain any observed similarities or differences.