COL 726 Homework 6

16–23 August, 2020

1. Prove that a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its restriction to any line is convex, i.e. for all $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ the function $\tilde{f}(t) = f(\mathbf{x} + t\mathbf{v})$ is convex.

   What if we restrict attention only to lines parallel to the coordinate axes? That is, suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function such that for all $\mathbf{x} \in \mathbb{R}^n$ and all $i = 1, \ldots, n$, the function $\tilde{f}(t) = f(\mathbf{x} + t\mathbf{e}_i)$ is convex. Is this sufficient for f to be convex? Prove or give a counterexample.

2. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form (note the absolute value signs!)

$$f(\mathbf{x}) = \sum_{i=1}^{k} |\mathbf{a}_i^T \mathbf{x} + b_i|.$$

   Explain why f is convex but cannot be minimized effectively with gradient descent or Newton's method. Then, prove that the smoothed function $\hat{f}(\mathbf{x}) = \sum_{i=1}^{k} \sqrt{(\mathbf{a}_i^T \mathbf{x} + b_i)^2 + \epsilon}$ is convex for any constant $\epsilon > 0$. For full marks, do so <u>without</u> calculating any Hessians.

3. You are given an undirected graph on $(m+n)$ nodes, each associated with a position $\mathbf{x}_i \in \mathbb{R}^k$. The positions of m nodes are fixed, while those of the remaining n are to be chosen to optimize the total cost (here $p > 1$ is a real number):[1]

$$\text{minimize} \sum_{(i,j)\in E} \|\mathbf{x}_i - \mathbf{x}_j\|_2^p.$$

   (a) Interpreting the optimization variables as a single vector $\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^{nk}$, find the gradient of the objective function.

   Implement a function to find the minimum using gradient descent with backtracking line search. Terminate when the norm of the gradient is below a user-specified threshold.

   (b) Then, implement a function to solve the problem using Newton's method, terminating when the Newton decrement is below a user-specified threshold.

   <u>Note:</u> If the Hessian is too complicated to derive, you may estimate it via centered differences applied to the gradient, then ensure symmetry via $\mathbf{H} \leftarrow \frac{1}{2}(\mathbf{H} + \mathbf{H}^T)$.

   Test your implementation on a graph of your choice, with $k = 2$, $m, n \geq 5$, and $p = 1.2, 2, 5$. Visualize the optimal configurations, similarly to B&V Figs. 8.15–17. Also report the objective value and gradient norm as a function of iteration number for both algorithms.

[1] This is the nonlinear facility location problem described in B&V Ch. 8.7, but you don't have to read that to solve this problem.