ECE 4803, Fall 2020

Homework #7

Due Tuesday, November 3, at 11:59pm

1. Prepare a one paragraph summary of what we talked about in class since the last assignment. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other classes?). The more insight you give, the better.

2. In the notes we showed that the subdifferential of $\|\boldsymbol{x}\|_1$ at $\boldsymbol{x}$ is given by vectors $\boldsymbol{u}$ that satisfy

$$\begin{aligned} u_n &= \operatorname{sign}(x_n) \quad && \text{if } x_n \neq 0, \\ |u_n| &\leq 1 && \text{if } x_n = 0. \end{aligned}$$

   Note that we could also write this as

$$\partial\|\boldsymbol{x}\|_1 = \left\{\boldsymbol{u} \,:\, \|\boldsymbol{u}\|_\infty = 1, \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|_1\right\}.$$

   It turns out that the subdifferential of $\|\boldsymbol{x}\|_\infty$ takes the related form:

$$\partial\|\boldsymbol{x}\|_\infty = \left\{\boldsymbol{u} \,:\, \|\boldsymbol{u}\|_1 = 1, \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|_\infty\right\}.$$

   (a) Describe a simple procedure for constructing a vector $\boldsymbol{u} \in \partial\|\boldsymbol{x}\|_\infty$ from $\boldsymbol{x}$.

   (b) Show that if $\boldsymbol{u}$ satisfies $\|\boldsymbol{u}\|_1 = 1$ and $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|_\infty$, then it must be a subgradient.

   (c) (Optional.) Provide an argument that if $\boldsymbol{u}$ is a subgradient of $\|\boldsymbol{x}\|_\infty$, then it must satisfy $\|\boldsymbol{u}\|_1 = 1$ and $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|_\infty$.

3. We have now spent quite a while looking at least squares problems where we aim to minimize $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$. While minimizing the ℓ_2 norm of the error is *often* a good idea, a big part of why it is so popular is just that it is easy to minimize. But now that we know about nonsmooth optimization, we can explore situations where other ℓ_p norms might be more appropriate. In particular, we will consider minimizing $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_1$ and $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty$ using the subgradient method.

   (a) In the notes we described how to construct a subgradient for $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_1$. Using a similar approach (and guided by the previous problem) show how to construct a subgradient for $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty$.

   (b) Download the file `hw07_prob3.py`. This file sets up a simple regression problem in which $\boldsymbol{b}$ consists of noisy observations of a smooth function $f(t)$ and considers three noise models: Gaussian noise, sparse Gaussian noise, and uniform noise. Compute and plot the least squares solution for each case.

   (c) Now implement subgradient descent for $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_1$. Use a backtracking line search to set the step size α_k. Note that there are two related implementation challenges to think about. First, you can pick *any* subgradient. This gives you some freedom whenever an entry of $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$ is equal to zero. Feel free to pick any rule you like here. The other consideration is that, unless you are lucky, your procedure for selecting a subgradient will not necessarily result in a subgradient of $\mathbf{0}$ at the solution. Thus, the norm of the subgradient is not a good test for convergence. Instead you can either check to see when the objective function stops improving or when the iterates $\boldsymbol{x}^{(k)}$ stop changing. (You should probably check this to make sure it is constant for several iterations in a row.) Compute and plot the resulting estimate of $\boldsymbol{x}$ for each of the three noise cases.

(d) Next repeat the implementation process from part (c) for $\|\boldsymbol{Ax} - \boldsymbol{b}\|_\infty$. Again, compute and plot the results for each of the three noise cases.

(e) Compare the results that you obtained in the previous parts. Which method seems best suited for each case?