ECE 6270, Spring 2021

Homework #5

Due Sunday, March 28, at 11:59pm Suggested Reading: B&V, Sections 3.3, 5.1, 5.2, and 5.5.

1. Prepare a one paragraph summary of what we talked about 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. Provide feedback to your peers on Homework #4 in Canvas.

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

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

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

4. **The support function and dual norms.** In this problem we will explore the notion of the support function and dual norms for some common ℓ_p norms. In the problems below, let $\mathcal{B}_p = \{\boldsymbol{x} \in \mathbb{R}^N : \|\boldsymbol{x}\|_p \leq 1\}$, where $\|\cdot\|_p$ denotes the standard ℓ_p norm on $\mathbb{R}^N$, and let $\|\cdot\|_{p*}$ denote the dual norm corresponding to $\|\cdot\|_p$.

   (a) Compute the support function $h_{\mathcal{B}_2}(\boldsymbol{\nu})$.

   (b) Show that the ℓ_2 norm is "self-dual": $\|\cdot\|_{2*} = \|\cdot\|_2$.

   (c) Compute the support function $h_{\mathcal{B}_1}(\boldsymbol{\nu})$.

   (d) Show that $\|\cdot\|_{1*} = \|\cdot\|_\infty$.

   (e) Compute the support function $h_{\mathcal{B}_\infty}(\boldsymbol{\nu})$.

   (f) Show that $\|\cdot\|_{\infty*} = \|\cdot\|_1$.

5. **The convex conjugate and subdifferentials.** There is a close relationship between the subdifferential of a function f and its convex conjugate f^*. In this problem you will explore some of these connections. In the problems below, we will assume that f is convex.

   (a) Show that if $\boldsymbol{\nu} \in \partial f(\boldsymbol{x})$, then $\boldsymbol{x} \in \partial f^*(\boldsymbol{\nu})$.

   (b) Now suppose that f closed (and hence $f^{**} = f$). Show that if $\boldsymbol{x} \in \partial f^*(\boldsymbol{\nu})$ then $\boldsymbol{\nu} \in \partial f(\boldsymbol{x})$.

6. **Descent cones.** In our discussions about the theory and practice of unconstrained optimization of differentiable functions, we often referred to the notion of a *descent direction*. Recall that $\boldsymbol{d}$ is a descent direction for f at a point $\boldsymbol{x}_0$ if

$$f(\boldsymbol{x}_0 + t\boldsymbol{d}) < f(\boldsymbol{x}_0) \quad \text{for some } t > 0.$$

   For convex, differentiable f, this is equivalent to the condition $\langle \boldsymbol{d}, \nabla f(\boldsymbol{x}_0) \rangle < 0$.

   (a) Prove that for a (not necessarily differentiable) convex function, the set

$$\mathcal{D}(\boldsymbol{x}_0) = \{\boldsymbol{d} : f(\boldsymbol{x}_0 + t\boldsymbol{d}) \leq f(\boldsymbol{x}_0) \text{ for some } \ t > 0\}$$

   is a convex cone. $\mathcal{D}(\boldsymbol{x}_0)$ is called the *descent cone* or *cone of descent* of f at $\boldsymbol{x}_0$.

   (b) Describe $\mathcal{D}(\boldsymbol{x}_0)$ for $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$.

   (c) Describe $\mathcal{D}(\boldsymbol{x}_0)$ for $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$.
   (Your answer should be a very clean expression in terms of the set Γ_0 where $\boldsymbol{x}_0$ is non-zero and the signs of the entries of $\boldsymbol{x}_0$ on Γ_0.)

7. **Lagrangian duality.** Here we explore Lagrangian duality with a simple example. Consider the optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}}{\text{minimize}} \ & x^2 + 1 \\ \text{subject to} \ & (x-2)(x-4) \leq 0. \end{aligned}$$

   (a) Provide as simple as possible of a description of the feasible set.

   (b) Determine both the minimizer $x^\star$ as well as the value of the objective function at the minimizer.

   (c) Plot the objective function, indicating in your plot the feasible set. Also plot the Lagrangian $\mathcal{L}(x, \lambda)$ for a few values of λ.

   (d) Derive and plot the dual function $d(\lambda)$.

   (e) State the dual problem and find the maximizer $\lambda^\star$ as well as the value $d(\lambda^\star)$. Does strong duality hold?

Last updated 15:38, March 21, 2021

8. **Penalization for equality constraints.** Consider the constrained optimization problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; & f(\boldsymbol{x}) \\ \text{subject to} \;\; & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{1}$$

where $f(\boldsymbol{x})$ is convex and differentiable and $\boldsymbol{A}$ is an $M \times N$ matrix. Next, consider the unconstrained problem given by

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; f(\boldsymbol{x}) + \alpha\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2, \tag{2}$$

where $\alpha > 0$ is a parameter. Intuition would suggest that the solution to (2) should be an approximation to the original problem. Let $\widetilde{\boldsymbol{x}}$ denote a solution to (2), and consider the vector defined by $\widetilde{\boldsymbol{\nu}} = 2\alpha(\boldsymbol{A}\widetilde{\boldsymbol{x}} - \boldsymbol{b})$. Show how to use $\widetilde{\boldsymbol{\nu}}$ to derive a lower bound on the optimal value of (1).

9. **Optimal power allocation for Gaussian channels.** There is a classic problem in information theory for which the KKT conditions reveal an elegant solution. The scenario is as follows: we are trying to communicate from point A to point B, and we have N different channels available. The channels are Gaussian in that the relationship between the input I_n and output O_n is

$$O_n = I_n + Z_n, \quad Z_n \sim \text{Normal}(0, \sigma_n^2).$$

It is a classic result in information theory that at most we can learn

$$C_n = \frac{1}{2}\log_2\left(1 + \frac{p_n}{\sigma_n^2}\right), \quad p_n = \mathrm{E}[I_n^2],$$

bits of information about I_n from the observation O_n per transmission — this quantity is called the *channel capacity*. We can interpret p_n as the (average) power needed to transmit I_n and p_n/σ_n^2 as the *signal to noise ratio*. As we commit more power to the transmission in channel n, we are able to communicate more information per usage. If we assign powers $p_1, \ldots, p_N$ to each of the channels, the total capacity is

$$C = \sum_{n=1}^{N} \frac{1}{2}\log_2\left(1 + \frac{p_n}{\sigma_n^2}\right).$$

Suppose now that we have a constraint on the total (expected) power we can use for a transmission:

$$\sum_{n=1}^{N} p_n \le p.$$

In this problem we will ask: what is the optimal way to allocate this power between the N channels to maximize the capacity?

We can cast this question as a convex optimization program. First, it is pretty clear that no matter what, we want to use the maximum power available to us, so we will just take the constraint above to be an equality constraint. Let x_n be the fraction of the total power allocated to channel n:

$$x_n = \frac{p_n}{p}.$$

Then

$$\log_2\left(1 + \frac{p_n}{\sigma_n^2}\right) = \log_2(\alpha_n + x_n) - \log_2(\alpha_n), \quad \alpha_n = \frac{\sigma_n^2}{p}.$$

Our task is then to solve

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{maximize}} \ \frac{1}{2}\sum_{n=1}^{N} \log_2(\alpha_n + x_n) - \log_2(\alpha_n) \quad \text{subject to} \quad x_n \geq 0, \quad \sum_{n=1}^{N} x_n = 1.$$

This is equivalent to

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ -\sum_{n=1}^{N} \log_2(\alpha_n + x_n) \quad \text{subject to} \quad \boldsymbol{x} \geq 0, \quad \mathbf{1}^{\mathrm{T}}\boldsymbol{x} = 1.$$

(a) Write down the KKT conditions for this problem. There are N inequality constraints and 1 equality constraint, so your answer should relate the optimal $\boldsymbol{x}^\star$ to $\boldsymbol{\lambda}^\star$ and $\nu^\star$.

(b) Simplify your answer to the previous part so that it only depends on $\nu^\star$.

(c) Show how given the single number $\nu^\star$ we can recover the optimal $\boldsymbol{x}^\star$.

(d) Write the equality constraint in terms of $\nu^\star$. Describe an algorithm for finding the $\nu^\star$ that meets this constraint.

(e) Make an informative sketch and interpret your result.

10. **The LASSO.** In this problem you will implement both subgradient descent and proximal gradient descent to solve the LASSO:

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \tau\|\boldsymbol{x}\|_1.$$

In the problems below, you will evaluate your code by testing it on the problem defined by the following code:

```
import numpy as np
np.random.seed(2021) # Set random seed so results are repeatable
# Set parameters
M = 100
N = 1000
S = 10

# Define A and y
A = np.random.randn(M,N)
```

```
ind0 = np.random.choice(N,S,0) # index subset
x0 = np.zeros(N)
x0[ind0] = np.random.rand(S)
y = A@x0 + .25*np.random.randn(M)
```

(a) Use CVXPY to solve the LASSO using the data above for a few values of τ. What value of τ seems to work best? Use this value in the subsequent parts of this problem.

(b) Implement subgradient descent for this problem. Produce a plot showing the value of the objective function as a function of iteration number. Show results for the following step size selection rules: $\alpha_k = \alpha$, $\alpha_k = \alpha/\sqrt{k}$, $\alpha_k = \alpha/k$. For each rule tune α to get reasonable performance.

(c) Implement the proximal gradient method for this problem (without acceleration). Use a fixed step size α. You may tune this manually, but there is also a principled choice. Produce a plot showing the value of the objective function as a function of iteration number

(d) Implement the proximal gradient method with acceleration. Use the same choice of α as in the previous part and use the rule $\beta_k = (k-1)/(k+2)$. Produce a plot showing the value of the objective function as a function of iteration number