Convex optimization exercises - GD, ctd.

Projections

The Projected Gradient Descent algorithm (PGD) requires computing ℓ_2 projections onto the feasible set. This is a difficult (convex optimization) problem in general, but in some cases it can be done very efficiently.

Exercise 1. *Show how to compute projection onto:*

- *unit* ℓ_2*-ball?*
- *unit* ℓ_∞*-ball?*
- *unit* ℓ_1*-ball?*

A lower bound

We will now prove that optimizing an M-smooth convex function by GD requires $\frac{cMR^2}{T^2}$ steps, for some c ($R = ||x^* - x_1||_2$). **Note:** This bound is actually asymptotically tight. And it can also be modified for M-smooth m-strongly convex case.
One possible "difficult function" is the following:

$$f_k(x_1, \ldots, x_k) = x_1^2 + \sum_{i=1}^{k-1} (x_i - x_{i+1})^2 + x_k^2 - 2x_1.$$

You can think about this function as describing some sort of energy of a set of k balls connected in a loop with strings. Here x_i describe positions (on a line) of balls, and one of the balls (x_0) is dropped from the formula because its position is fixed to be 0. Moreover, the first ball is being pulled to the right. Minimizing this energy function corresponds to determining the equilibrium positions of the balls. We claim that GD does that very poorly, hitting the bound stated in previous paragraph.

1. Show that f_k is convex.
2. Compute the gradient and Hessian of f_k.
3. Show that f_k is 8-smooth (i.e. eigenvalues of the Hessian are ≤ 8).
4. Find the minimizer of f_k.
5. Show that after T steps of GD starting at $x_1 = 0$, we have $x_T = x_{T+1} = \ldots = x_k = 0$. What is the best solution satisfying these constraints?
6. To prove the bound we use $k = 2T + 1$.
7. What is the optimality gap in this case?
8. What is R^2 in this case?
9. Show the bound stated at the start of the exercise (using $(M/8)f_k$).