2 Projecting onto the L1-ball [Kevin, 25 points]

In this problem we will devise an efficient algorithm for projecting a vector $y \in \mathbb{R}^n$ onto the unit $L1$-ball. That is, to solve the optimization problem

$$x^* = \arg\min_x \; \|x - y\|_2^2/2, \quad \text{subject to:} \tag{1}$$

$$\|x\|_1 \leq 1. \tag{2}$$

(a) [2 points] Write out the Legrangian, $L(x, \lambda)$, where $\lambda \in \mathbb{R}$.

This problem is convex and strictly feasible, so by Slater's condition strong duality holds and therefore the KKT conditions are necessary. Convexity implies the KKT conditions are sufficient.

(b) [4 points] Write out the KKT conditions that an optimal primal/dual pair, (x^*, λ^*), must satisfy.

(c) [3 points] Show that if $\|y\|_1 \leq 1$ then $x^* = y, \lambda^* = 0$ satisfy the KKT conditions.

Assume from this point that $\|y\|_1 > 1$. Otherwise, by (c), we know x^*.

(d) [4 points] Prove by contradiction and the KKT conditions that $\lambda^* > 0$ and $\|x^*\|_1 = 1$.

For any $\lambda \geq 0$, let us define $x(\lambda) = S_\lambda(y)$. That is,

$$x_i(\lambda) = \begin{cases} y_i - \lambda & \text{if } y_i \geq \lambda \\ y_i + \lambda & \text{if } y_i \leq -\lambda \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

(e) [5 points] Prove that $(x(\lambda), \lambda)$ satisfy the stationarity condition (*i.e.*, first-order optimality). That is, $(x(\lambda), \lambda)$ satisfy all the KKT conditions, but perhaps not primal feasibility.

Note: $(x(\lambda), \lambda)$ satisfy the stationarity condition uniquely.

Let $f(\lambda) = \|x(\lambda)\|_1$. We've now shown that if we find a $\lambda^* \geq 0$ such that $f(\lambda^*) = 1$, then $x(\lambda^*)$ satisfies the KKT conditions and solves our problem.

(f) [2 points] Show that $f(0) > 1$ and there is a λ^+ such that $f(\lambda^+) = 0$.

It is apparent that $f(\lambda)$ is continuous (it is also decreasing and convex). Therefore, by the intermediate value theorem, we have that λ^* exists. Now let's quickly compute λ^* exactly.

It is not hard to see that λ^* is invariant under permutations of the elements of y, or changes to their signs. Let's consider the vector $\tilde{y}$, which has the same elements as $|y|$, sorted in descending order. That is, $\tilde{y}_1 \geq \tilde{y}_2 \geq \ldots \geq \tilde{y}_n \geq 0$. Given an index $k \in \{1, 2, \ldots, n-1\}$, we know that choosing $\lambda \in [\tilde{y}_k, \tilde{y}_{k+1}]$ will make all $\tilde{x}_j(\lambda) = 0$, for $j \in \{k+1, k+2, \ldots, n\}$.

(g) [5 points] Using no more than $\Sigma_k = \sum_{i=1}^{k} \tilde{y}_i$, and the values $\tilde{y}_k, \tilde{y}_{k+1}$ give a simple formula for $\lambda^* \in [\tilde{y}_k, \tilde{y}_{k+1}]$ if it exists, or a test that concludes λ^* is not in that range.

That is, to project $\|y\|_1 > 1$ onto the $L1$-ball, we start by sorting, $\tilde{y} = \text{sort}(|y|)$, then find λ^* by testing each interval with part (h), and finish by computing $x^* = S_{\lambda^*}(y)$ by soft-thresholding. This all takes only $O(n \log n)$ time.

(By the way, a similar procedure can be derived for projecting onto the probability simplex—another frequently-encountered set.)