CoCoA @ Amazon

Simone Forte
ETH Zurich
fortesi@student.ethz.ch

We wish to solve the following minimization problem:

$$\kappa\|\alpha\|_1 + \frac{1}{2n}\|A\alpha - b\|_2^2$$

First of all, we'll rewrite it a form compatible to the one studied in CoCoA+:

$$-\frac{1}{n}\sum_{i=1}^{n} n\kappa|\alpha_i| - \frac{\lambda}{2}\|A\alpha - b\|_2^2$$

with $\lambda = \frac{1}{n}$ and we'll define: $\ell(\alpha_i) = n\kappa|\alpha_i|$. We now notice that for an optimal solution α to this problem, it holds that for every i:

$$\ell(\alpha_i) \leq \|b\|_2^2 := B$$

since the solution with $\alpha = \mathbf{0}$ has objective value B. We can therefore define an $\bar{\ell}$ to be as follows:

$$\bar{\ell}(\alpha_i) = \begin{cases} \ell(\alpha_i) & : \ell(\alpha_i) \leq B \\ \ell(\alpha_i) + \big(\ell(\alpha_i) - B\big)^2 & : otherwise \end{cases}$$

and the following problem will thus clearly have the same optimal solution as the one defined with ℓ:

$$D(\alpha) = -\frac{1}{n}\sum_{i=1}^{n} \bar{\ell}_i(\alpha_i) - \frac{\lambda}{2}\|A\alpha - b\|_2^2.$$

We'll now compute the dual conjugate of $\bar{\ell}$, which is as follows (a proof will come):

$$\bar{\ell}^*(x) = \begin{cases} 0 & : \frac{x}{n\kappa} \in [-1, 1] \\ \frac{1}{4}\left(\frac{x}{n\kappa}\right)^2 + \frac{(2B-1)}{2}\frac{x}{n\kappa} - \frac{4B-1}{4} & : otherwise \end{cases}$$

The convenient thing about this conjugate with respect to the indicator function (the conjugate of ℓ) is of course that is defined on the entire $\mathbb{R}$. Another thing that is possible to show is that, being $n\kappa|\alpha_i| \leq B$ we have that $|\alpha_i| \leq \frac{B}{n\kappa}$. Using this we can bound the maximum $|w^T A_i|$ as follows:

$$|w^T A_i| = |(A\alpha - b)^T A_i| \leq |A\alpha^T A_i| + |b^T A_i| \leq \sum_{j=1}^{n} |\alpha_j||A_j^T A_i| + |b^T A_i| \leq$$

$$\frac{B}{n\kappa} n \max_{i=1..n} \|A_i\|^2 + \max_{i=1..n} |b^T A_i| = \frac{B}{\kappa} \max_{i=1..n} \|A_i\|^2 + \max_{i=1..n} |b^T A_i| := C$$

This bound is important since this gives us that, for every w explored by the optimization algorithm, $x = w^T A_i$ is in $[-C, C]$ and the function, being a quadratic, is therefore $\Theta(C)$-Liptschitz in this interval. We should therefore be able to employ Theorem 8 from the CoCoA+ paper!

References