Deflected Subgradient Convergence Analysis

Elia Piccoli
Nicola Gugole

1 Recap

First of all a quick recap of the different aspects of the analysis.
Current formulation of the problem

$$
\begin{aligned}
\max_{\beta_i} & -\frac{1}{2}\sum_i \sum_j \beta_i \beta_j K(x_i, x_j) \\
& -\epsilon \sum_i |\beta_i| \\
& +\sum_i y_i \beta_i \\
& With\ the\ constraints \qquad \begin{cases} \sum_i \beta_i = 0 \\ \beta_i \in [-C,\ C] \end{cases}
\end{aligned}
\tag{1}
$$

Direction projection formulation

$$
\begin{aligned}
\min_{s} \quad & \frac{1}{2}\left\|d - s\right\|^2 \\
& With\ the\ constraints \qquad \left\{ (\beta_i - s_i) \in [-C,\ C] \right.
\end{aligned}
\tag{2}
$$

β projection formulation

$$
\begin{aligned}
\min_{\beta_{proj}} \quad & \frac{1}{2}\left\|\beta - \beta_{proj}\right\|^2 \\
& With\ the\ constraints \qquad \begin{cases} \sum_i \beta^i_{proj} = 0 \\ \beta^i_{proj} \in [-C,\ C] \end{cases}
\end{aligned}
\tag{3}
$$

2 Projection Algorithms

In this section the focus will be on how the two projection problems are solved.

- (3) $\longrightarrow$ *Convex Separable Knapsack Problem algorithm.* This projection is *easy* to perform and can also be attained fast with algorithms in the family of $\mathcal{O}(n \cdot log(n))$ / $\mathcal{O}(n)$, as stated in [see 3, Introduction] and more in depth analyzed in [see 1, Algorithm 3.1 (convergence proved in Remark 3.2(d))]

- (2) $\longrightarrow$ *Projected Gradient method for BCQP*. This projection is even *easier* to achieve wrt the previous one and can be performed using the same algorithm seen in class [see 4, Slide 13]. This will converge to the global minimum given the structure of problem (2) with a convergence rate of $\mathcal{O}(\frac{1}{\epsilon})$.

3 Deflected Subgradient Algorithm

The approach that we are going to analyze is a *Constrained Deflected Subgradient Method* using *Target Value Stepsize* with a *Non-Vanishing Threshold*. With this method f^* is approximated by an estimate that is updated as the algorithm proceeds. The estimate will be the *target level* which is defined wrt two values: f^k_{ref} which is the *reference value*, and δ_k which is the *threshold*. This two values will be used to approximate f^* in the formulation of the stepsize. In particular the stepsize has to follow a constraint between the α and ψ parameter (*stepsize restriction*) to assure convergence.

$$0 \leq \nu_k = \psi_k \frac{f_k - f^k_{ref} + \delta_k}{\|d_k\|^2} \qquad 0 \leq \psi_k \leq \alpha_k \leq 1 \tag{4}$$

Giving a general algorithm for solving (1):

Algorithm 1: Deflected Subgradient Algorithm
(variable x stands for β)

```
begin
    $xref \longleftarrow x$
    $fref \longleftarrow inf$
    $\delta \longleftarrow 0$
    $dprev \longleftarrow 0$
    while $true$ do
        $v \longleftarrow \frac{1}{2}x'Qx + qx$
        $g \longleftarrow Qx + q$
        Check if in $stopped$ condition
        Check if in $optimal$ condition
        // reset $\delta$ if $v$ is $good$ or decrease it otherwise
        if $v \leq fref - \delta$ then
            $\delta \longleftarrow \delta\_reset \cdot \max v, 1$
        else
            $\delta \longleftarrow \max(\delta\rho, eps \cdot \max(|\min(v, fref)|, 1))$
        end
        // update $fref$ and $xref$ if needed
        if $v < fref$ then
            $fref \longleftarrow v$
            $xref \longleftarrow x$
        end
        $d \longleftarrow \alpha g + (1 - \alpha)dprev$
        $d \longleftarrow Project(d)$                // project $d$ solving (2)
        $dprev \longleftarrow d$
        $\lambda \longleftarrow v - fref + \delta$
        $\nu \longleftarrow \frac{\psi \cdot \lambda}{\|d\|^2}$          // stepsize-restricted $\rightarrow \psi \leq \alpha$
        $x \longleftarrow x - \nu \cdot d$
        $x \longleftarrow Project(x)$                // project $x$ solving (3)
    end
end
```

The projections required in Algorithm 1 are the ones presented in Section 2. The two projections are *easy* to perform, allowing the convergence of the *Deflected Subgradient Algorithm* as stated in [see 2, Theorem 3.6]. The theorem has two conditions to ensure the convergence:

- [see 2, Cond 2.13] is satisfied since both the direction used in the current iteration and the direction used in the next iteration are the projection of the deflected direction of the current iteration [see *Deflected Subgradient Algorithm*].

- [see 2, Cond 3.5] is satisfied since the λ is always greater or equal to zero because of the algorithm structure [see *Deflected Subgradient Algorithm*].

In conclusion, since the requirements are satisfied the algorithm converges. The convergence rate expected is at best the convergence rate of a SM using *Polyak* (since the algorithm proposed is an approximation using *Target Level*), suggesting a best convergence of $\mathcal{O}(\frac{1}{\epsilon^2})$ [see 5, Slide 41] .

References

[1] Krzysztof C. Kiwiel. “Breakpoint searching algorithms for the continuous quadratic knapsack problem”. In: *Mathematical Programming* 112.2 (Apr. 2008), pp. 473–491. ISSN: 1436-4646. DOI: 10.1007/s10107-006-0050-z. URL: https://doi.org/10.1007/s10107-006-0050-z.

[2] d’Antonio Giacomo and Frangioni Antonio. “Convergence Analysis of Deflected Conditional Approximate Subgradient Methods”. In: *SIUM Journal on Optimization* 20 (Jan. 2009). DOI: 10.1137/080718814.

[3] Frangioni Antonio and Enrico Gorgone. “A library for continuous convex separable quadratic knapsack problems”. In: *European Journal of Operational Research* 229 (Aug. 2013), pp. 37–40. DOI: 10.1016/j.ejor.2013.02.038.

[4] Frangioni Antonio. *Constrained optimization*. URL: https://elearning.di.unipi.it/pluginfile.php/43257/mod_resource/content/2/7-constrained_optimization.pdf.

[5] Frangioni Antonio. *Unconstrained optimization III Less-than-gradient methods*. URL: https://elearning.di.unipi.it/pluginfile.php/42987/mod_resource/content/2/5-unconstrained%20optimization%20III.pdf.