

1 Projection formulation

Starting from the following formalization:

$$\begin{aligned} \min_{\beta_i} \quad & \frac{1}{2} \sum_i \sum_j \beta_i \beta_j K(x_i, x_j) \\ & + \varepsilon \sum_i |\beta_i| \\ & - \sum_i y_i \beta_i \end{aligned} \tag{1}$$

$$\text{With the constraints} \quad \begin{cases} \sum_i \beta_i = 0 \\ \beta_i \in [-C, C] \end{cases}$$

The optimization problem can be solved by the following steps, including the deflection as specified by the task:

Algorithm 1: Compute β

```

1 while optima not found do
2    $v \leftarrow f(\beta)$ 
3    $g \leftarrow \nabla f(\beta)$ 
4   check if current solution is best so far (to update)
5    $d \leftarrow \gamma g + (1 - \gamma) d_{prev}$ 
6    $stepsize \leftarrow \frac{\psi(v - f_{ref} + \delta)}{\|d\|^2}$ 
7    $\beta \leftarrow \beta - stepsize \cdot d$ 
8   project  $\beta$  by solving the KP
9 end
```

Projection of line 8 is where a simple formulation can be exploited to solve efficiently a Convex Separable Knapsack Problem. The main objective of the formulation is keeping the resulting β_{proj} vector as similar as possible to the one produced by line 7 while at the same time producing a β_{proj} vector satisfying the constraints. Therefore:

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

Which by lagrangian relaxation leads to:

$$\min_{\beta_{proj}} \frac{1}{2} \sum_i ((\beta^i - \beta_{proj}^i)^2 + \mu \beta_{proj}^i) \quad (3)$$

With the constraints $\left\{ \beta_{proj}^i \in [-C, C] \right\}$

2 Alternative formulation

A different concept came up while discussing with each other, probably slightly more intensive on the computational side (because of the more complex constraints), but carrying the same underlying thought process. Instead of focusing on ensuring the constraints exploiting as reference the β vector from line 7 of [1] we could ensure them by exploiting the d vector from line 5. This would imply projecting the deflected direction trying to keep as small as possible the variation to it, which in the language of mathematics results in:

$$\min_s \frac{1}{2} \|d - s\|^2 \quad (4)$$

Obviously the constraints are different in this occasion. The only constraints are on β and since the supposedly new β is constructed using the chosen direction (i.e. $\beta = \beta - \text{stepsize} \cdot s$) we can test and ensure the constraints by deriving the ones relative to s :

$$\left\{ \begin{array}{l} s_i \in [\frac{\beta_i - C}{\text{stepsize}}, \frac{\beta_i + C}{\text{stepsize}}] \\ \sum_i s_i = \frac{\sum_i \beta_i}{\text{stepsize}} \end{array} \right. \quad (5)$$

Which, wrapping up by lagrangian relaxation, leads to:

$$\min_s \frac{1}{2} \sum_i (d_i - s_i)^2 + \mu (s_i - \frac{\sum_j \beta_j}{\text{stepsize}}) \quad (6)$$

With the constraints $\left\{ s_i \in [\frac{\beta_i - C}{\text{stepsize}}, \frac{\beta_i + C}{\text{stepsize}}] \right\}$

stepsize is obtained using original direction d so to avoid a recursive constraint on s . This is not a dramatic approximation since the objective of the minimization of (4) is keeping s as close as possible to d .

3 Wrapping Up

The two formulations are both separable problems with constraints of the knapsack type and the underlying mathematical meaning seems accurate in both the cases.

Obviously the age-old question remains: are we still missing on something? Is one of the two more correct than the other?