

## Notes on the weights in Hierbasis

According to (Zhao et al. 2009). Consider the regularized estimates given by

$$\hat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}} \{L(Z, \beta) + \lambda \cdot T(\beta)\}$$

where  $Z = (X, Y)$  is the observed data.

We consider CAP (Composite Absolute Penalties) family of penalties. They are highly customizable and build upon  $L_\gamma$  penalties to express both grouped and hierarchical selection. CAP penalties are convex whenever all norms used in its construction are convex. \*\* While the CAP estimates are not sparser than LASSO, they result in more parsimonious use of degrees of freedom and more stable estimates.

Let each node correspond to a group of variables  $G_k$  and set its descendants to be the groups that should only be added to the model after  $G_k$ . CAP penalties enforcing the hierarchy can be obtained by setting

$$T(\beta) = \sum_{m=1}^{\text{nodes}} \alpha_m \cdot \|(\beta_{G_m}, \beta_{\text{all descendants of } G_m})\|_{\gamma_m},$$

with  $\alpha_m > 0$  for all  $m$ . The factor  $\alpha_m$  can be used to correct for the effect of a coefficient being present in too many groups

Theorem: If  $\gamma_i \geq 1$ , then  $T(\beta)$  is convex. If, in addition, the loss function  $L$  is convex in  $\beta$ , the objective function of the CAP optimization problem is convex.

## Notes on Group Lasso

According to (Yuan and Lin 2006). In regression problems, we want to find important explanatory factors in predicting the response variable, where each explanatory factor may be represented by a group of derived inputs. For example, in the additive model with polynomial or nonparametric components (each component in the additive model may be expressed as a linear combination of a number of basis functions of the original measured variable). Variable selection amounts to selection of a group of variables instead of individual variables (ie selecting one basis expansion).

For a vector  $\eta \in \mathbb{R}^d$ ,  $d \geq 1$ , and a symmetric  $d \times d$  positive definite matrix  $K$ , we denote

$$\|\eta\|_K = (\eta^T K \eta)^{1/2}$$

(write  $\|\eta\| = \|\eta\|_{I_d}$ ). Given the positive definite matrices  $K_1, \dots, K_J$ , the group lasso estimate is the solution to

$$\frac{1}{2} \|Y - \sum_{j=1}^J X_j \beta_j\|^2 + \lambda \sum_{j=1}^J \|\beta_j\|_{K_j}.$$