A primal-dual algorithm for group ℓ_1 regularization with overlapping groups

Appendix

Proof of Lemma 1

One of the advantages of Algorithm 2 for computing the projection $\pi_{\tau K}$, is the fact that the constraints that are already satisfied can be discarded. This follows from Lemma 1. To prove it we need an additional lemma.

Lemma 2. *Let $\beta \in \mathbb{R}^d$. Then $|\pi_{\tau K}(\beta)|_i \leq |\beta|_i$ for all $i = 1, \ldots, d$.*

Proof. *Let $C_r = \{\beta \in \mathbb{R}^d, \|\beta\|_{G_r} \leq 1\}$, and define the sequence of projections π_n by setting $\pi_n = \pi_{\tau C_{n \bmod B}}$. Then, given a sequence $\alpha_n \in (0,1)$ converging to 0, denote by $\beta_{n+1} = \alpha_n \beta + (1-\alpha_n)\pi_n(\beta^n)$, with $\beta^0 := \beta$.*

By Theorem 3.1 in [4] it follows that for every $\beta \in \mathbb{R}^d$

$$\pi_{\tau K}(\beta) = \lim_n \beta^n, \tag{1}$$

if α_n goes to zero not too fast. We now prove by induction that (1) implies $|(\beta^n)_j| \leq |\beta_j|$ for every n and for $j = 1, \ldots, d$. First observe that, since we are projecting on a cylinder with radius τ and centered on a coordinate axis (or coordinate subspace)

$$\begin{aligned} |\beta_j^1| &= |\alpha_1 \beta_j + (1-\alpha_1)(\pi_{C_1}(\beta^0))_j| \\ &\leq \alpha_1 |\beta_j| + (1-\alpha_1)|\beta_j| \\ &= |\beta_j|. \end{aligned}$$

Now suppose the inequality $|(\beta^n)_j| \leq |\beta_j|$ to be satisfied. We have to prove the same for β^{n+1}. By definition

$$\begin{aligned} |\beta_j^{n+1}| &= |\alpha_n \beta_j + (1-\alpha_n)(\pi_n(\beta^n))_j| \\ &\leq \alpha_n |\beta_j| + (1-\alpha_n)|(\pi_n(\beta^n))_j| \\ &\leq |\beta_j|. \end{aligned}$$

Passing to the limit we finally get:

$$|(\pi_{\tau K}(\beta))_j| = \lim_n |(\beta_n)_j| \leq |\beta_j|.$$

If we denote $K^{-s} = \bigcap_{r \neq s} C_r$, the above Lemma guarantees in particular that if $\pi_{\tau K^{-s}}(\beta) \in \tau K$ for all $\beta \in \tau C_s$, then $\pi_{\tau K^{-s}}(\beta) = \pi_{\tau K}(\beta)$. Lemma 1 then follows.

Proof of Theorem 2

As usual, the Lagrangian function for the minimization problem (4) is defined as

$$
\begin{aligned}
L(v,\lambda) &= \|v-\beta\|^2 + \sum_{r=1}^{\hat{B}} \lambda_r(\|v\|_{G_r}^2 - \tau^2) \\
&= \sum_{j=1}^{d}\left[(v_j-\beta_j)^2 + \sum_{r=1}^{\hat{B}} \lambda_r \mathbf{1}_{r,j} v_j^2\right] - \sum_{r=1}^{\hat{B}} \lambda_r \tau^2 \\
&= \sum_{j=1}^{d}(1+\sum_{r=1}^{\hat{B}}\mathbf{1}_{r,j}\lambda_r)\left(v_j - \frac{\beta_j}{1+\sum_{r=1}^{\hat{B}}\mathbf{1}_{r,j}\lambda_r}\right)^2 - \sum_{j=1}^{d}\frac{\beta_j^2}{1+\sum_{r=1}^{\hat{B}}\mathbf{1}_{r,j}\lambda_r} - \sum_{r=1}^{\hat{B}}\lambda_r\tau^2 + \|\beta\|^2
\end{aligned} \tag{2}
$$

where $\lambda \in \mathbb{R}^{\hat{B}}$, and $\mathbf{1}_{r,j}$ is 1 if j belongs to group G_r and 0 otherwise. The dual function is then

$$
g(\lambda) = \inf_{v\in\mathbb{R}^d} L(v,\lambda) = L\left(\frac{\beta_j}{1+\sum_{r=1}^{\hat{B}}\mathbf{1}_{r,j}\lambda_r}, \lambda\right) = -\sum_{j=1}^{d}\frac{\beta_j^2}{1+\sum_{r=1}^{\hat{B}}\mathbf{1}_{r,j}\lambda_r} - \sum_{r=1}^{\hat{B}}\lambda_r\tau^2 + \|\beta\|^2 .
$$

Since strong duality holds, the minimum of (4) is equal to maximum of the dual problem which is therefore

$$
\begin{array}{ll}
\text{Maximize} & g(\lambda) \\
\text{subject to} & \lambda_r \geq 0 \text{ for } r = 1,\ldots,\hat{B}.
\end{array} \tag{3}
$$

Once the solution λ^* to the dual problem (3) is obtained, the solution to the primal problem (4), v^*, is given by

$$
v_j^* = \frac{\beta_j}{1+\sum_{r=1}^{\hat{B}}\lambda_r^*\mathbf{1}_{r,j}} \qquad \text{for } j = 1,\ldots,d.
$$

The dual problem (3) can thus be efficiently solved via Algorithm 2, where the first and second partial derivative of $g(\lambda)$ are given by

$$
\partial_r g(\lambda) = \tau^2 - \sum_{j=1}^{d}\beta_j^2\frac{\mathbf{1}_{r,j}}{(1+\sum_{s=1}^{\hat{B}}\mathbf{1}_{s,j}\lambda_s)^2},
$$

and

$$
\partial_r\partial_s g(\lambda) = \sum_{j=1}^{d}\frac{2\beta_j^2\mathbf{1}_{r,j}\mathbf{1}_{r',j}}{(1+\sum_{s=1}^{\hat{B}}\mathbf{1}_{s,j}\lambda_s)^3} = \begin{cases} 0 & \text{if } \hat{G}_r \cap \hat{G}_{r'} = \emptyset \\ 2\sum_{j\in\hat{\mathcal{G}}_{r,r'}}\beta_j^2(1+\sum_{s=1}^{\hat{B}}\mathbf{1}_{s,j}\lambda_s)^{-3} & \text{otherwise.} \end{cases}
$$