# A family of penalty functions for structured sparsity

Charles A. Micchelli, Jean M. Morales (speaker) and Massimiliano Pontil

NIPS 2010

December 8th, 2010

# Sparsity with Lasso and variational form

We consider a linear regression model under **structured sparsity** assumptions on $\beta^*$, the underlying vector. A popular estimate for the sparse case is

$$\hat{\beta}^{\mathrm{LASSO}} = \operatorname*{argmin}_{\beta} \left\{ \|y - X\beta\|_2^2 + \|\beta\|_1 \right\}.$$

Equivalently, the $\ell_1$ norm is the solution of a **convex** problem:

$$\|\beta\|_1 = \inf_{\lambda > 0} \frac{1}{2} \sum_{i=1}^{n} \left( \frac{\beta_i^2}{\lambda_i} + \lambda_i \right),$$

where $\lambda \in \mathbb{R}^n$ is a vector of **auxiliary** variables.

# Constrained variational form

We introduce the **additional constraint** $\lambda \in \Lambda \subseteq \mathbb{R}^n_{++}$:

$$\Omega(\beta|\Lambda) = \inf_{\lambda \in \Lambda} \frac{1}{2} \sum_{i=1}^{n} \left( \frac{\beta_i^2}{\lambda_i} + \lambda_i \right).$$

The set $\Lambda$ provides a **structure** in the sense that the function is minimum for $|\beta|$ in the set:

$$\Omega(\beta|\Lambda) = \|\beta\|_1 \Leftrightarrow |\beta| \in \Lambda.$$

If the set $\Lambda$ is convex, so is function $\Omega$, and we have a **convex optimization problem**.

# Examples of structured sparsity

Several possibilities for the set $\Lambda$ has been investigated.

- The graph with incidence matrix $A$:

$$\Lambda = \left\{ \lambda \in \mathbb{R}^n_{++} : A\lambda \geq 0 \right\}.$$

- Using $D^k$, the $k$-th order difference operator:

$$\Lambda = \left\{ \lambda \in \mathbb{R}^n_{++} : D^k(\lambda) \geq 0 \right\}.$$

These and other sets constraint directly $\lambda$, and indirectly $|\beta|$.