Optimality conditions for the LASSO

Given a $y \in \mathbb{R}^M$ and a $M \times N$ matrix Φ, the LASSO solves

$$\min_x \; \frac{1}{2}\|\Phi x - y\|_2^2 + \tau\|x\|_1. \tag{1}$$

To make things easier (i.e. smooth), introduce the auxiliary variable t, and rewrite (1) as

$$\min_{x,t} \frac{1}{2}\|\Phi x - y\|_2^2 + \tau \cdot \mathbf{1}^T t \quad \text{subject to} \quad x - t \le 0, \quad -x - t \le 0. \tag{2}$$

(By $\mathbf{1}$ we mean an N-vector of all ones.) The Lagrange function for (2) is

$$L(x, t, \lambda_+, \lambda_-) = \frac{1}{2}\|\Phi x - y\|_2^2 + \tau \cdot \mathbf{1}^T t + {\lambda_+}^T(x - t) + {\lambda_-}^T(-x - t),$$

and so solving (2) is equivalent to solving

$$\min_{x,t} \max_{\lambda_+,\lambda_-} L(x, t, \lambda_+, \lambda_-), \quad \text{subject to} \quad \lambda_-, \lambda_+ \ge 0,$$

and the dual program is

$$\max_{\lambda_+,\lambda_-} \left(\min_{x,t} L(x, t, \lambda_+, \lambda_-) \right) \quad \text{subject to} \quad \lambda_-, \lambda_+ \ge 0.$$

For fixed λ_+, λ_-, we minimize L over x and t by finding where the gradient is zero (L is convex in x and t). We have

$$\nabla_{x,t} L = \begin{bmatrix} \Phi^T(\Phi x - y) + \lambda_+ - \lambda_- \\ \tau \cdot \mathbf{1} - \lambda_+ - \lambda_- \end{bmatrix}.$$

Set $v = -(\lambda_+ - \lambda_-)$. For $\nabla_{x,t} L = 0$, we need

$$\begin{aligned} -\lambda_+ + \lambda_- &= \Phi^T(\Phi x - y), \\ \lambda_+ + \lambda_- &= \tau \cdot \mathbf{1}. \end{aligned} \tag{3}$$

Since $\lambda_+, \lambda_- \ge 0$, we can combine these two constraints to get the dual feasibility condition

$$\|\Phi^T(\Phi x - y)\|_\infty \le \tau. \tag{4}$$

Now, note that in light of (3), we can re-write the Lagrangian as

$$\begin{aligned} L(x, t, \lambda_+, \lambda_-) &= \frac{1}{2}\|\Phi x - y\|_2^2 + \tau \mathbf{1}^T t + \lambda_+(x - t) + \lambda_-(-x - t) \\ &= \frac{1}{2}\|\Phi x - y\|_2^2 + \tau \mathbf{1}^T t + (\lambda_+ - \lambda_-)^T x + (-\lambda_+ - \lambda_-)^T(-x - t) \\ &= \frac{1}{2}\|\Phi x - y\|_2^2 + (y - \Phi x)^T \Phi x, \end{aligned}$$

which now depends only on x. At the solution $x^\star$, the Lagrangian will match the primal value,

$$L(x^\star) = \frac{1}{2}\|\Phi x^\star - y\|_2^2 + \tau\|x^\star\|_1$$

and so $x^\star$ must obey

$$(y - \Phi x)^T \Phi x = \tau \|x\|_1.$$

Since we have the dual feasibility condition (4), this means that we must have

$$\begin{aligned} \Phi_\Gamma^T (y - \Phi x^\star) &= \tau \,\mathrm{sgn}(x_\Gamma^\star) \\ \|\Phi_{\Gamma^c}^T (y - \Phi x^\star)\|_\infty &\leq \tau \end{aligned}$$

where Γ is the set of locations on which $x^\star$ is supported, Φ_Γ are the columns from Φ indexed by Γ, and $x_\Gamma^\star$ is the $|\Gamma|$-vector containing the non-zero coefficients of $x^\star$.