Zerrad
Avigail
(MASH)

DM 3

1) We rewrite the optimization problem as following:

$$\min_{v,w} \frac{1}{2}\|v\|_2^2 + \lambda\|w\|_1$$
$$\text{s.t } v = Xw - y$$

Lagrangian: $L(w,v,\mu) = \frac{1}{2}\|v\|_2^2 + \lambda\|w\|_1 + \mu^T(Xw - y - v)$

Lagrange dual function: $g(\mu) = \inf_{w,v} L(w,v,\mu)$

$$= \inf_{w,v} \frac{1}{2}\|v\|_2^2 + \lambda\|w\|_1 + \mu^T(Xw - y - v)$$

$$= \inf_{w} (\lambda\|w\|_1 + \mu^T Xw) + \inf_{v} \left(\frac{1}{2}\|v\|_2^2 - \mu^T v\right) - \mu^T y$$

↳ using the definition of the conjugate function to the ℓ_1-norm, we have:

$$\inf_{w} (\lambda\|w\|_1 + \mu^T Xw) = \begin{cases} 0 & \text{if } \|X^T\mu\|_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases}$$

↳ $\inf_{v} \left(\frac{1}{2}\|v\|_2^2 - \mu^T v\right)$. By setting the gradient at 0, we have: $v - \mu = 0 \Leftrightarrow v = \mu$

$$\Rightarrow \inf_{v} \left(\frac{1}{2}\|v\|_2^2 - \mu^T v\right) = \frac{1}{2}\|\mu\|_2^2 - \|\mu\|_2^2 = -\frac{1}{2}\|\mu\|_2^2$$

Then, $g(\mu) = \begin{cases} -\frac{1}{2}\|\mu\|_2^2 - \mu^T y & \text{if } \|X^T\mu\|_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases}$

Dual problem: $\max -\frac{1}{2}\|\mu\|_2^2 - y^T\mu$

$$\text{s.t } \|X^T\mu\|_\infty \leq \lambda$$

We can rewrite the condition $\|X^T\mu\|_\infty \leq \lambda$ as $\begin{pmatrix} X^T \\ -X^T \end{pmatrix}\mu \leq \lambda \mathbb{1}_{2d}$,

and then the dual problem of LASSO is:

$$\min \ \mu^T Q \mu + y^T \mu \quad \text{s.t. } A\mu \leq b \qquad \text{with } Q = \tfrac{1}{2} I_n, \quad A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}, \quad b = \lambda \mathbb{1}_{2d}$$

2) we want to solve the problem $\min \ v^T Q v + p^T v$

s.t $Av - b \leq 0$

There is no equality constraints; the associated centering problem is:

$$\min_v \ t(v^T Q v + p^T v) - \sum_{i=1}^{2d} \log(b_i - A_i v)$$

where A_i is the i-th row of A.

We denote $\phi(v) := -\sum_{i=1}^{2d} \log(b_i - A_i v)$, and $g_t(v) = t(v^T Q v + p^T v) + \phi(v)$

$$\nabla\phi(v) = -\sum_{i=1}^{2d} \frac{-A_i^T}{b_i - A_i v} \quad \Rightarrow \quad \nabla g_t(v) = t((Q+Q^T)v + p) + \sum_{i=1}^{2d} \frac{A_i^T}{b_i - A_i v}$$

$$\nabla^2\phi(v) = \sum_{i=1}^{2d} \frac{A_i^T A_i}{(A_i v - b_i)^2} \quad \Rightarrow \quad \nabla^2 g_t(v) = t(Q+Q^T) + \sum_{i=1}^{2d} \frac{A_i^T A_i}{(A_i v - b_i)^2}$$

We use Newton's method for unconstrained optimization to solve $\min_v g_t(v)$.

3) We observe that when μ is small ($\mu = 2$), we do small centering steps and it takes a lot of Newton iterations to get to a solution with precision ε. For higher values of μ, we do longer steps on the central path and the total number of Newton iterations is much less than in the case with a small μ. We also notice that the number of Newton iterations for $\mu \geq 10$ stays fairly constant when we increase μ. Thus, it's sufficient to choose $\mu = 10$.

Here is the figure obtained for a precision $\varepsilon=10^{-6}$, for different values of μ.