Matthis
Maillard

Convex Optimization - Homework 3

1) The LASSO problem is: minimize $\frac{1}{2}\|Xw - y\|_2^2 + \lambda\|w\|_1$
with $w \in \mathbb{R}^d$, $X \in \mathbb{R}^{n \times d}$, $y \in \mathbb{R}^n$ and $\lambda > 0$

We introduce a new variable: $z \in \mathbb{R}^n$ and $z = Xw$

The problem becomes:

$$\text{minimize } \frac{1}{2}\|z - y\|_2^2 + \lambda\|w\|_1$$
$$\text{subject to } Xw = z$$

The Lagrangian of the problem is:

$\mathcal{L}(w, z, v) = \frac{1}{2}\|z - y\|_2^2 + \lambda\|w\|_1 + v^T(z - Xw)$ with $v \in \mathbb{R}^n$

The Lagrangian is streactly convex with respect to z. Thus, to minimize it, we set its gradient to zero.

$\nabla_z \mathcal{L}(w, z, v) = 0 \Leftrightarrow \cancel{z} - y + v = 0 \Leftrightarrow z = y - v$

We have seen in homework 2 that $\sup_x (y^T x - \|x\|_1) = \begin{cases} 0 & \text{if } -1 \preceq y \preceq 1 \\ +\infty & \text{otherwise} \end{cases}$

Hence, $\inf_w (\lambda\|w\|_1 - v^T X w) = -\sup_w (-\lambda\|w\|_1 + v^T X w) = \begin{cases} 0 & \text{if } -\lambda \preceq X^T v \preceq \lambda \\ -\infty & \text{otherwise} \end{cases}$

So $\inf_{z,w} \mathcal{L}(w, z, v) = -\frac{1}{2} v^T v + v^T y$ if $-\lambda \preceq X^T v \preceq \lambda$

The dual of the LASSO problem is:

maximize $-\frac{1}{2} v^T v + v^T y$

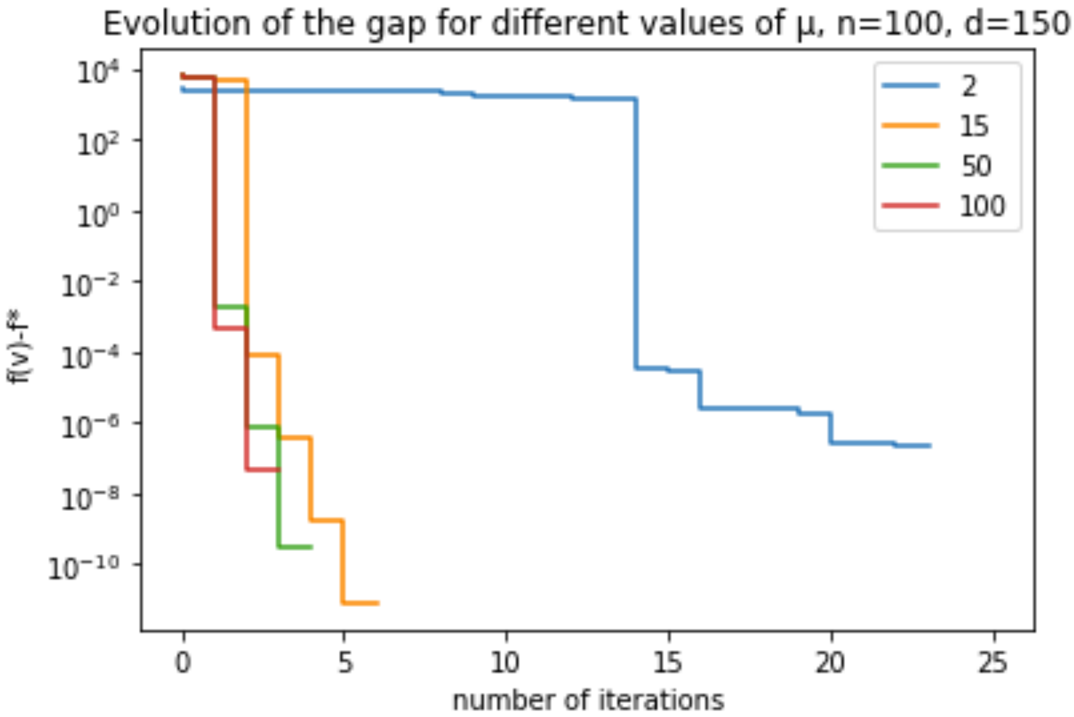
subject to $-\lambda \preceq X^T v \preceq \lambda$

which is equivalent to:

minimize $\frac{1}{2} v^T v - y^T v$

subject to $Av \preceq \lambda$ with $A = \begin{pmatrix} X^T \\ -X^T \end{pmatrix}$

3)
The graphic was obtained with a precision criterion: epsilon = 10^-5



This are the best values that are obtained for the same problem, with different μ.

μ	2	15	50	100
f*	-6909.036379252145	-6909.0363792522	-6909.036379251936	-6909.036379252208

We can see that μ does not really impact the value of f*. According to 1), we have Xw = z = v - y. X and y are fixed, so if μ has an impact on v, it has an impact on w. Since, the value of f* does not depend on μ, we can conclude that μ does not impact v therefore not on w also. μ has an impact on the number of iterations and the convergence time. If we increase μ, the number of iterations in the centering step will decrease, but the number of iterations in the centering step will increase. So we have to choose a correct size for μ, not too big but not too small. As there are only 2 iterations more with μ=15 than with 50 or 100, we will choose this one. We could not have chosen 2 because the number of iteration in the barrier step is too high.