Rendu DM3 Optimisation Convexe

Yoann Pradat

November 12, 2018

1. Let's derive the dual of the LASSO problem in the variable $w \in \mathbf{R}^d$.

$$\min_{w \in \mathbf{R}^d} \frac{1}{2}\|Xw - y\|_2^2 + \lambda \|w\|_1 \tag{1}$$

where $\lambda \in \mathbf{R}$, $X \in \mathbf{R}^{n \times d}$ and $y \in \mathbf{R}^n$ are fixed. The LASSO problem is equivalent to the constrained problem over $(v, w) \in \mathbf{R}^{n \times d}$

$$\begin{aligned} &\min_{(v,w) \in \mathbf{R}^{n \times d}} \frac{1}{2}\|v\|_2^2 + \lambda \|w\|_1 \\ &\text{s.t} \quad Xw - y = v \end{aligned} \tag{2}$$

Let $f_0 : (v, w) \mapsto f_1(v) + f_2(w)$ be the objective function with $f_1 : v \mapsto \frac{1}{2}\|v\|_2^2$ and $f_2 : w \mapsto \lambda\|w\|_1$. Let's compute the conjugage of f_0, it will be useful in the formulation of the dual problem. First notice that f_1 and f_2 are convex and that f_0 is a function of two variables but that these variables are separated. Consequently, the conjugate of f_0 is the sum of the conjugates of f_1 and f_2 i.e $f_0^*(s, t) = f_1^*(t) + f_2^*(s)$.

The conjugate of f_1 is $f_1^*(t) = \frac{1}{2}\|t\|_{*,2}^2$ where $\|t\|_{*,2} = \sup\{t^T v | \|v\|_2 \leq 1\}$. Indeed let $t \in \mathbf{R}^n$. Then

$$\begin{aligned} &\forall v \in \mathbf{R}^n \quad t^T v \leq \|t\|_{*,2}\|v\|_2 \\ &\quad \implies t^T v - \frac{1}{2}\|v\|_2^2 \leq \|t\|_{*,2}\|v\|_2 - \frac{1}{2}\|v\|_2^2 \\ &\quad \implies t^T v - \frac{1}{2}\|v\|_2^2 \leq \frac{1}{2}\|t\|_{*,2}^2 \end{aligned}$$

as the left-hand side on the second line is a quadratic function of $\|v\|_2$ with maximum $\frac{1}{2}\|t\|_{*,2}^2$. Therefore, $f_1^*(t) \leq \frac{1}{2}\|t\|_{*,2}^2$. Now, let $v \in \mathbf{R}^n$ be such that $t^T v = \|t\|_{*,2}\|v\|_2$. We can scale v so that $\|v\|_2 = \|t\|_{*,2}$. Then $t^T v - \frac{1}{2}\|v\|_2^2 = \frac{1}{2}\|t\|_{*,2}^2$. As a consequence $f_1^*(t) \geq \frac{1}{2}\|t\|_{*,2}^2$ and eventually $f_1^*(t) = \frac{1}{2}\|t\|_{*,2}^2$.

The conjugate of f_2 is

$$f_2^*(s) = \begin{cases} 0 & \|\lambda s\|_{*,1} \leq 1 \\ \infty & \text{otherwise} \end{cases}$$

Let $s \in \mathbf{R}^d$. If $\|\lambda s\|_{*,1} > 1$, by definition of the dual norm there exists $w \in \mathbf{R}^d$ with $\|w\|_1 \leq 1$ and $\lambda s^T w > 1$. Taking $z = aw$ and letting $a \to \infty$ we have

$$\lambda s^T z - \|z\|_1 = a(\lambda s^T w - \|w\|_1) \to \infty$$

If $\|\lambda s\|_{*,1} \leq 1$, it is easy to show that $\lambda s^T w - \|w\|_1$ is bounded above by 0 and 0 is attained for $w = 0$.

Let's now compute the dual of the problem (2). For $\nu \in \mathbf{R}^n$ we have

$$g(\nu) = \inf_{(v,w)\in\mathbf{R}^{n\times d}} \Big(f_1(v) + f_2(w) + \nu^T(Xw - y - v)\Big)$$
$$= -f_1^*(\nu) - f_2^*(-X^T\nu) - \nu^T y$$

Given the previous calculations, ν is dual feasible if and only if $\|\lambda X^T \nu\|_{*,1} \leq 1$. The dual norm of the l_2-norm is the l_2-norm and the dual norm of the l_1-norm is the l_∞-norm. Therefore, the dual of problem (2) is

$$\begin{aligned} &\max_{\nu\in\mathbf{R}^n} -\frac{1}{2}\|\nu\|_2^2 - \nu^T y \\ &\text{s.t} \quad \|\lambda X^T\nu\|_\infty \leq 1 \end{aligned} \tag{3}$$

that is to say

$$\begin{aligned} &\min_{v\in\mathbf{R}^n} \frac{1}{2}v^T v + v^T y \\ &\text{s.t} \quad X^T v \preceq \frac{1}{\lambda} \end{aligned} \tag{4}$$

The dual problem is thus obtained as a QP problem

$$\begin{aligned} &\min_{v\in\mathbf{R}^n} v^T Q v + v^T p \\ &\text{s.t} \quad A^T v \preceq b \end{aligned} \tag{5}$$

where $\boxed{Q = \frac{1}{2}I_n,\ p = y,\ A = X^T \text{ and } b = \frac{1}{\lambda}}$.

2. We are now going to solve the dual problem (5) using a barrier method. Given $(v_0, t_0, \mu, \epsilon)$

Barrier method
Repeat

1. Centering step. Compute $v^*(t)$ by minimizing $f(v) = \Big[t(v^T Q v + v^T p) - \sum_{i=1}^d \log(b_i - a_i^T v)\Big]$ with a_i^T is the i^{th} line of A.
2. Update $v := v^*(t)$
3. Stopping criterion stop if $m/t < \epsilon$
4. Increase t. $t := \mu t$.

The centering step is going to be solved using Newton method with backtracking line search that is to say

Centering step
Repeat

1. $\Delta v_{nt} := -\nabla^2 f(v)^{-1}\nabla f(v)$.
2. $\lambda^2 = \nabla f(v)^T \nabla^2 f(v)^{-1} \nabla f(v)$
3. Quit if $\frac{\lambda^2}{2} < \epsilon$
4. Choose step size s by backtrack line search.

5. Update $v := v + s\Delta v_{nt}$.

For fixed $t \in \mathbf{R}_{++}$, the gradient and the hessian of f are

$$\nabla f(v) = t(2Qv + p) + \sum_{i=1}^{d} \frac{a_i}{b_i - a_i^T v}$$

$$\nabla^2 f(v) = 2tQ + \left[\sum_{i=1}^{d} \frac{a_{ik} a_{il}}{(b_i - a_i^T v)^2}\right]_{k,l \in [\![1,n]\!]^2}$$

For given $\alpha \in (0, \frac{1}{2}), \beta \in (0, 1)$, the backtrack search is solved by

Backtrack search

1. Start at $s = 1$
2. Repeat $s := \beta s$ until $f(v + s\Delta v) < f(v) + \alpha s \nabla f(v)^T \Delta v$.

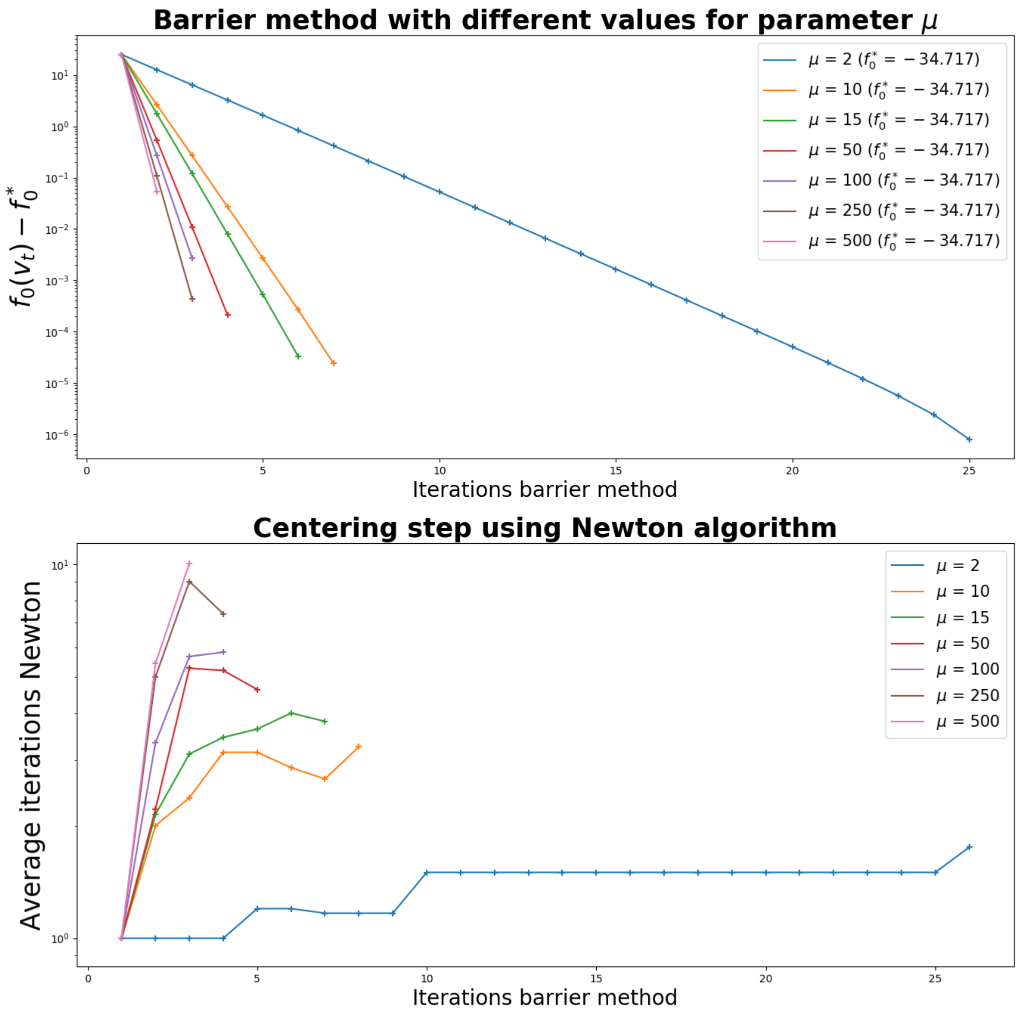

Figure 1 – Results barrier method on the dual of LASSO for random X and y, $\lambda = 10$

The constrained formulation of LASSO (2) is a convex problem for which Slater condition is satisfied. Therefore we have strong duality. Solving the dual problem, we have a dual feasible ν^* point for (5). Then, if (v^*, w^*) is the unique solution of

$$\min_{(v,w)} \Big(f_1(v) + f_2(w) + (\nu^*)^T (Xw - y - v) \Big) \tag{6}$$

and (v^*, w^*) is primal feasible, it is primal optimal.

The l_1-norm is not differentiable on any $w \in \mathbf{R}^d$ that has at least 1 null coordinate but is differentiable everywhere else. It is not obvious how to derive w after having solved the dual problem ...