Rep_4_LASSO

Szymon Czop

20 05 2020

Report LASSO optimization problem

We will start from exercises at the end of script from lecture.

Exercise 3 (page 12)

Let $A \in \mathbb{R}^{nxp}$ with $||A_1||_1 = \dots = ||A_n||_1 = 1$ and $b \in \mathbb{R}^n$ where $||A^Tb||_\infty > 0$. Let f, g be fuctions defined as follows:

$$\forall \lambda \in (0, ||A^Tb||_\infty], f(\lambda) = \lambda\Big(1 + \frac{||b||_2}{||A^Tb||_\infty}\Big) \; and \; g(\lambda) = -\frac{||A^Tb||_\infty ||b||_2}{\lambda} + ||b||_2 + ||A^Tb||_\infty$$

1)

Show that : $f(\frac{||A^Tb||_\infty}{2}) \leq 0 \; and \; g(\frac{||A^Tb||_\infty}{2} \leq 0)$

At the beginning notice that $||A^Tb||_\infty \leq ||b||_2$ because columns of A are normalized.

Proof:

$$f(\frac{||A^Tb||_\infty}{2}) = \frac{||A^Tb||_\infty}{2}\Big(1 + \frac{||b||_2}{||A^Tb||_\infty}\Big) - ||b||_2 = \frac{||A^Tb||_\infty}{2} + \frac{||b||_2}{2} - ||b||_2 = \frac{||A^Tb||_\infty - ||b||_2}{2} \leq 0$$

$$g(\frac{||A^Tb||_\infty}{2}) = -2\frac{||A^Tb||_\infty ||b||_2}{||A^Tb||_\infty} + ||b||_2 + ||A^Tb||_\infty = -2||b||_2 + ||b||_2 + ||A^Tb||_\infty = ||A^Tb||_\infty - ||b||_2 \leq 0 \quad \blacksquare$$

Knowing this we can say that safe and DPP rules do not discard columns of A when $\lambda \leq \frac{||A^Tb||_\infty}{2}$

2)

We will show that $f(||A^Tb||_\infty) = g(||A^Tb||_\infty) \; and \; f(\lambda) = 0$ if and only if $g(\lambda) = 0$

$f(||A^Tb||_\infty) = ||A^Tb||_\infty\Big(1 + \frac{||b||_2}{||A^Tb||_\infty}\Big) - ||b||_2 = ||A^Tb||_\infty + ||b||_2 - ||b||_2 = \frac{||A^Tb||_\infty ||b||_2}{||A^Tb||_\infty} + ||b||_2 + ||A^Tb||_\infty = g(||A^Tb||_\infty)$

Next

$$f(\lambda) = 0 \Leftrightarrow g(\lambda) = 0$$

$(\Rightarrow)$

$f(\lambda) = 0 \Rightarrow \lambda\Big(1 + \frac{||b||_2}{||A^Tb||_\infty}\Big) - ||b||_2 = 0 \Rightarrow \lambda = \frac{||A^Tb||_\infty ||b||_2}{||A^Tb||_\infty + ||b||_2} \Rightarrow g(\lambda) = g(\frac{||A^Tb||_\infty ||b||_2}{||A^Tb||_\infty + ||b||_2}) = -(||A^Tb||_\infty + ||b||_2) + ||b||_2 + ||A^Tb||_\infty = 0$

$(\Leftarrow)$

$$g(\lambda) = 0 \Rightarrow -\frac{||A^Tb||_\infty ||b||_2}{\lambda} + ||b||_2 + ||A^Tb||_\infty = 0 \Rightarrow \lambda = \frac{||A^Tb||_\infty ||b||_2}{||A^Tb||_\infty + ||b||_2}$$

$$\Rightarrow f(\lambda) = f(\frac{||A^Tb||_\infty ||b||_2}{||A^Tb||_\infty + ||b||_2}) = \frac{||b||_2(|A^Tb||_\infty + ||b||_2)}{||A^Tb||_\infty + ||b||_2} - ||b||_2 = 0\blacksquare$$

By proving this we see that both functions value at the end of λ field are the same. What is more certain λ has to make function g and f zero at the same time. If column A_j is discarded then both safe rule and DPP are doing it.

3)

We are going to prove that $g(|A_j^Tb|) \leq |A_j^Tb| \; and \; f(|A_j^Tb|) \leq |A_j^Tb|$.

$$f(|A_j^Tb|) = |A_j^Tb|\Big(1 + \frac{||b||_2}{||A^Tb||_\infty}\Big) - ||b||_2 \leq |A_j^Tb|\Big(1 + \frac{||b||_2}{|A_j^Tb|}\Big) - ||b||_2 = |A_j^Tb|$$

Now we will change a little our eqation: $\Big(\frac{||A^Tb||_\infty}{|A_j^Tb|} - 1\Big)(||b||_2 - |A_j^Tb|) \geq 0 \Leftrightarrow \frac{||A^Tb||_\infty ||b||_2}{|A_j^Tb|} - ||b||_2 - ||A^Tb||_\infty + |A_j^Tb| \geq 0 \Leftrightarrow -\frac{||A^Tb||_\infty ||b||_2}{|A_j^Tb|} + ||b||_2 + ||A^Tb||_\infty \leq |A_j^Tb| \Leftrightarrow g(|A_j^Tb|) \leq |A_j^Tb| \quad \blacksquare$

Having this we can conclude that safe rule and DPP do not discard column of A when $\lambda < |A_j^Tb|$

Exercises LAB 4

Safe rule and DPP rule

1)

I'm creating A matrix and b response vector as in exercise

To calcualte value of λ for each $x_i(\lambda) = 0$ where $i \in \{1, \dots, 4\}$. We are going to use both safe rule and DDP. They are desribed as follow:

x_j^* is zero for safe rule when :

$$|A_j^Tb| < \lambda\Big(1 + \frac{||A_j||_2||b||_2}{||A^Tb||_\infty}\Big) - ||A_j||_2||b||_2$$

x_j^* is zero DPP when:

$$|A_j^Tb| < -\frac{||A^Tb||_\infty ||A_j||_2|b||_2}{\lambda} + ||A_j||_2||b||_2 + ||A^Tb||_\infty$$

To calculate λ for safe rule we will change a little bit eqation to get direct value of lambda, same for DPP.

Changed safe rule:

$$\frac{|A_j^T b| + ||A_j||_2||b||_2}{\left(1 + \frac{||A_j||_2||b||_2}{||A^T b||_\infty}\right)} < \lambda$$

Changed DPP:

$$\frac{||A^T b||_\infty ||A_j||_2 |b||_2}{||A_j||_2||b||_2 + ||A^T b||_\infty - |A_j^T b|} < \lambda$$

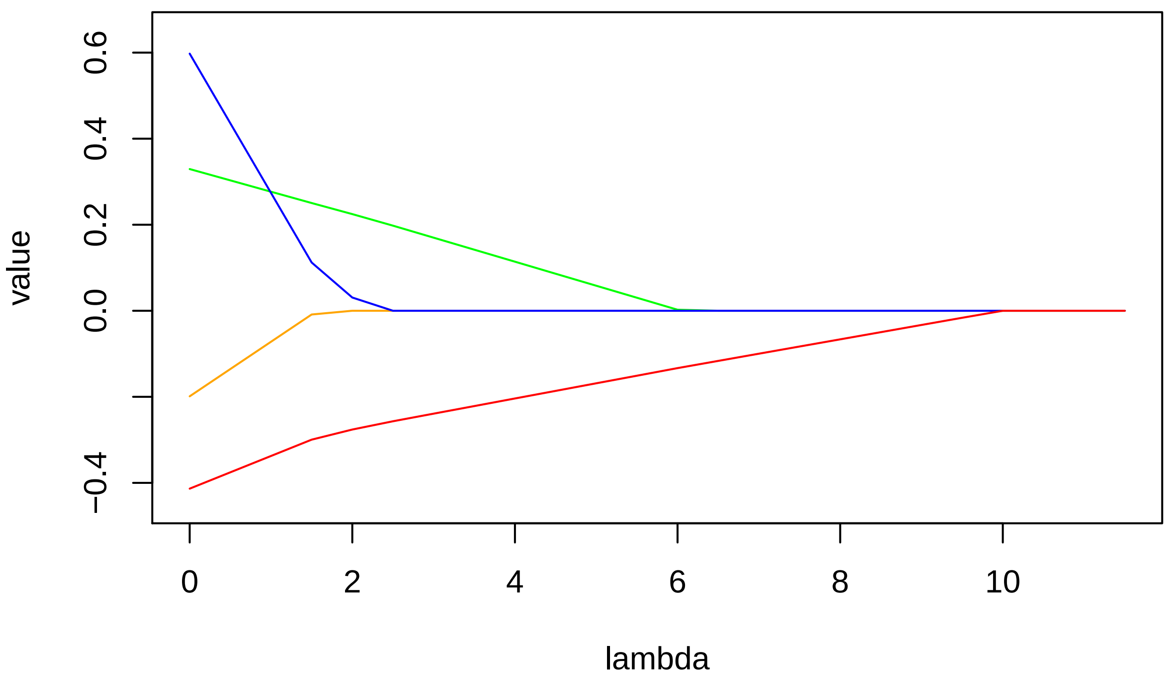
for all exaples $||A^T b||_\infty = 10$ and $||b||_2 = 5.5$.

Table 1: Lambda value

x_i	safe_rule	DPP
x_1	8.02	7.96
x_2	8.77	8.52
x_3	6.79	6.51
x_4	10.00	10.00

The values of λ under which given $x_i = 0$ are slightly different for both measures but only by tiny bit. Last value is the same because is bigger than real $||A^T b||_\infty$ and as we know if labda is bigger than this , we have zero vector.

2)



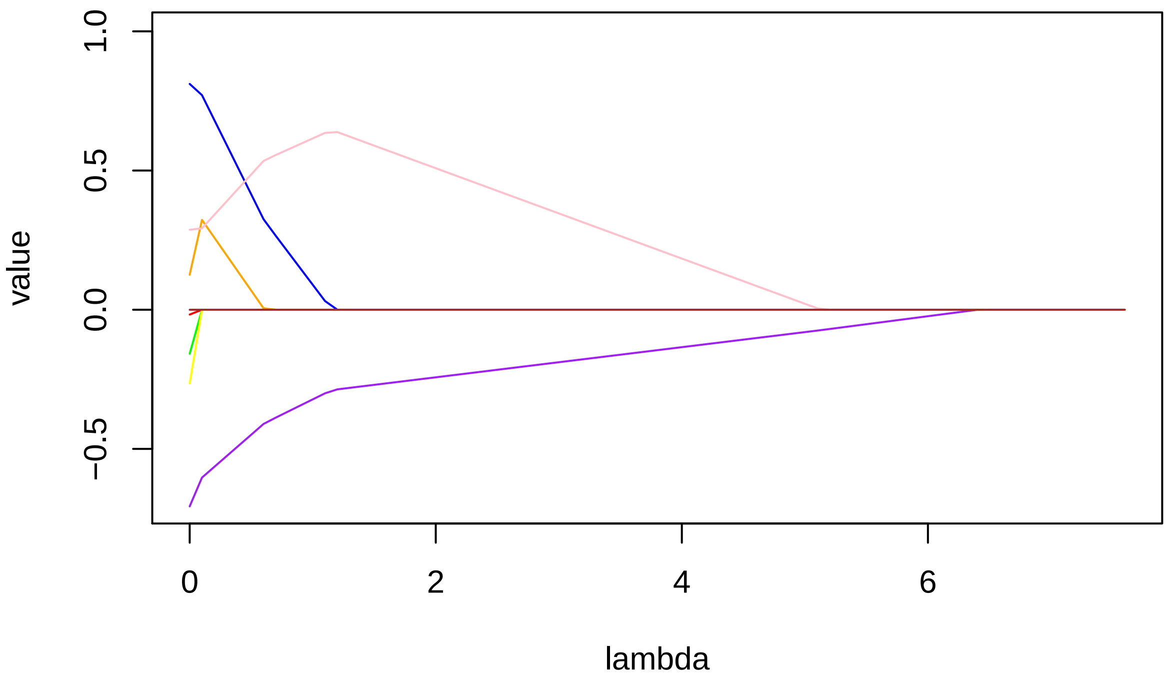
All the values of x (which is solution to LASSO problem) start from non-zero value. All lines are linear by parts. When λ is close to $||A^T b||_\infty$ all entries of x are becoming zero. Of course the bigger λ the more zero etries we have.

Exercise 2 (LASSO solution path)

Now we are changing dimesion of A matrix. Now $A \in \mathbb{R}^{4x8}$ so we have twice many columns as rows. This also has influence on b which is now $b \in \mathbb{R}^4$

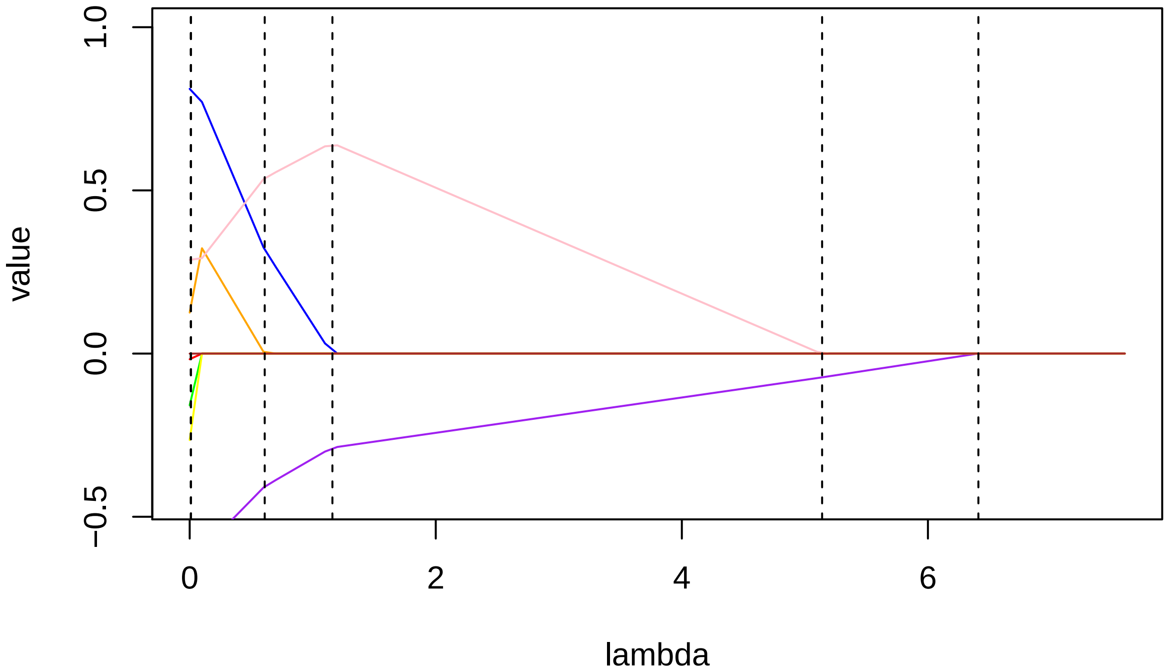
Plot is showing the function:

$$\lambda \in (0, 1.2||A^T b||_\infty] \rightarrow x_i(\lambda) \; where \; i \in \{1, 2, 3, 4\}$$



Again the lasso path start from non-zero values for each entry of x. Some of the values zero iteself very quickly and when $\lambda = 2$ We have only 2 non-zero entries left. Again when λ is close to $||A^T b||_\infty$, x is becoming a zero vector.

Now we wiil add vertical lines that will show where (with some ϵ) are lambdas that make coeffcients of x equal to zero:

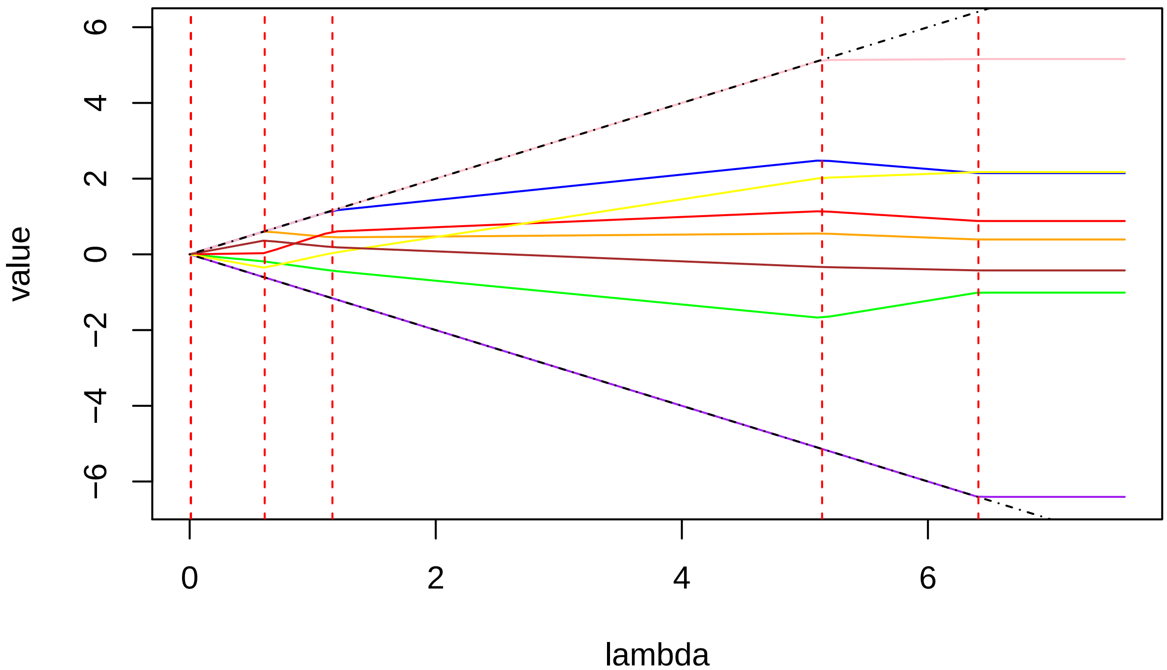


As we see dashed verical black lines always cross horizontal axis when one of the x_i is equal to zero. For calculation I made step eqal to 0.1 and this is first value where for my simulations $\lambda \neq 0$. As we see on the very left side dashed line is a bit behind point where some of x commpoents are zero. It means that for even smaller λ (in my case 0.01) these x_i would be zero as well. Again once $\lambda > ||A^T b||_\infty$ *then* $x(\lambda) = 0$. In this case $||A^T b||_\infty = 6.4$

2)

Now we will draw 3 functions in the same graph, and add verical lines from 1).

1) $\lambda \in (0, 1.2||A^T b||_\infty] \rightarrow A_i^T(b - Ax(\lambda))$ *where* $i \in \{1, 2, 3, 4\}$
2) $\lambda \in (0, 1.2||A^T b||_\infty] \rightarrow \lambda$
3) $\lambda \in (0, 1.2||A^T b||_\infty] \rightarrow -\lambda$

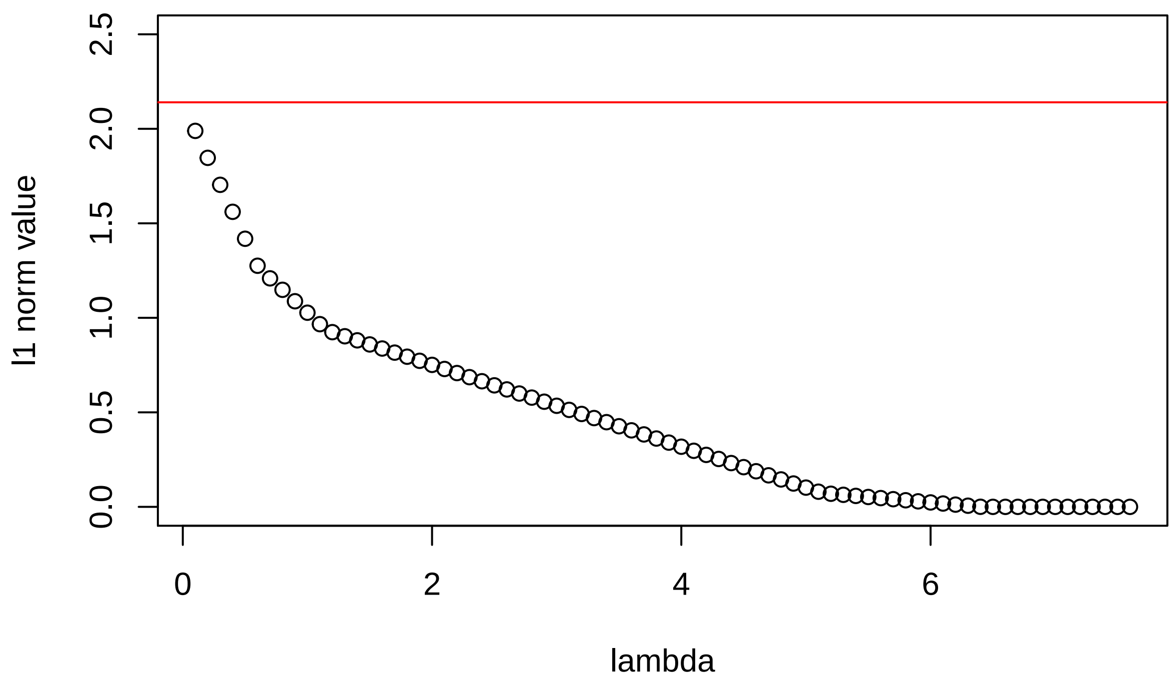


Verical red-dashed lines are crossing our lines in points where one of them is bending and start to linarly lower or increase(only at the beggining) it's value. Function 2) and 3) (black-dashed) are upper and lower boundry for the funcions and close them in a kind of "triangle". When $\lambda > ||A^T b||_\infty$ all the functions of x_i are becoming a constant lines and there will be no change for that because $x = 0$ vector.

3)

Let x^* be the basic pursuit solution of $Ax = b$

We are going to check graphically that $||x(\lambda)||_1 \leq ||x^*||_1$ and then prove it.



Horizontal red line is value of $||x^*||_1$ dot points are values of $||x(\lambda)||_1$ for different λ values. I got rid of value for $\lambda = 0$ because then solutions are close to x^* and this is not a matter of our concern in this point.

Now written proof:

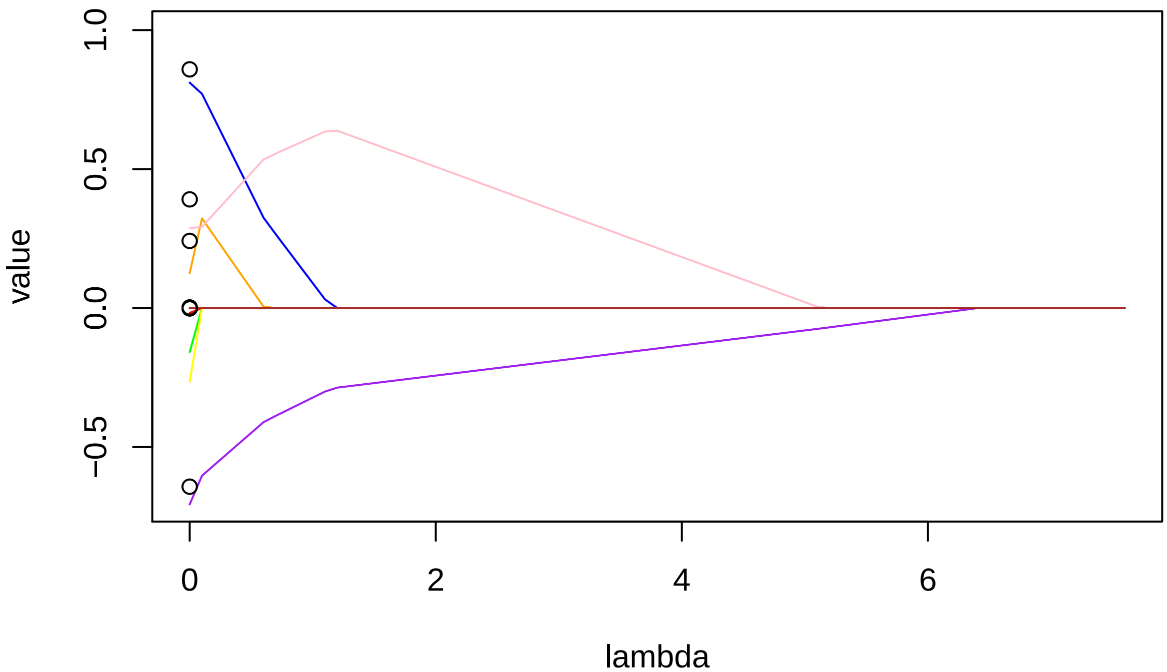
x^* is BP minimizer $x(\lambda)$ is lasso minimizer for given λ

$$\lambda||x(\lambda)|| \leq \frac{1}{2}||Ax(\lambda) - b||^2 + \lambda||x(\lambda)|| \leq \frac{1}{2}||Ax^* - b||^2 + \lambda||x^*|| = \lambda||x^*||_1$$

First inequality occurs because we simply add non negative value. Second because $x(\lambda)$ is lasso minimizer for give λ so if we use x^* this function is not minimized and thus higher. Last important fact is that $\frac{1}{2}||Ax^* - b||^2 = 0$ because x^* is BP minimizer.

4)

Now we will draw x^* value on the lasso path that we seen before. It will be done at $\lambda = 0$



There is visible discontinuity between curves and points at $\lambda = 0$ Points are moved a little bit moved up. By setting $\lambda = 0$ we have ordinary OLS problem thus outcome is different for BP minimizer.

Made and described by Szymon Czop