Hand-In 1 - FRTN50 @ LTH

Filip Kronström & Jonatan Kronander

September 2019

Task 1

Derive f^*:

$$f(x) = \frac{1}{2}x^TQx + q^Tx \tag{1}$$

$$\nabla f(x) = Qx + q \tag{2}$$

We have the definition for conjugate function in equation 3.

$$f^*(x) = \sup_x(s^Tx - f(x)) \tag{3}$$

Fermats rule states that x is the solution for sup if equation 4 is correct.

$$0 = s - \nabla f(x) = s - Qx + q \iff x = Q^{-1}(s - q) \tag{4}$$

If we apply this into the definition for f^* we get equations 5 and 6.

$$f^*(s) = s^TQ^{-1}(s - q) - \frac{1}{2}(s - q)^TQ^{-1}QQ^{-1}(s - q) - q^TQ^{-1}(s - q) \tag{5}$$

$$f^*(s) = (s^T - q^T)Q^{-1}(s - q) - \frac{1}{2}(s - q)^TQ^{-1}(s - q) = \frac{1}{2}(s - q)^TQ^{-1}(s - q) \tag{6}$$

Derive i_S^*:
We have the definition for conjugate function and we know that i_S is ∞ for every value X outside set S, hence following equations 7, 8, 9 and 10 is correct.

$$i_S^*(s) = \sup_X(s^Tx - i_S) = \sup_{X \in S}(s^Tx) \tag{7}$$

$$S = \{x : \forall_i, a_i \leq x_i \leq b_i\} \tag{8}$$

$$i_S^*(s) = s^Tx \tag{9}$$

$$x_i = \begin{cases} a_i & \text{if} \quad s_i < 0 \\ b_i & \text{if} \quad s_i \geq 0 \end{cases} \tag{10}$$

Write down the Fenchel-dual problem:

$$\min_{\mu}(f^*(\mu) + g^*(-\mu)) \tag{11}$$

Task 2

Show that f are L-smooth and find L:

We know that if f are L-smooth then we know that equation 12 is correct.

$$LI \succ \nabla^2 f = Q \iff 0 \succ LI - Q \tag{12}$$

Hence term $LI - Q$ must be positive definite and f is L-smooth if L is greater than each of every eigenvalue of Q. Else $LI - Q$ would have a negative eigenvalue and by definition not be positive definite.

Show that f^* are L^*-smooth and find L^*:

Similarly, f^* are L-smooth if and only if equation 13 is correct.

$$LI \succ \nabla^2 f^* = Q^{-1} \iff 0 \succ LI - Q^{-1} \tag{13}$$

Hence term $LI - Q^{-1}$ must be positive definite and f is L-smooth if L is greater than each of every eigenvalue of Q^{-1}.

Task 3

Derive expression for ∇f:

$$\nabla f = Qx + q \tag{14}$$

Derive expression for ∇f^*:

$$\nabla f^* = Q^{-1}(s - q) \tag{15}$$

Derive expression for $\mathbf{prox}_{\gamma i_S}$.

Fermat's rule implicates directly equation 16:

$$\text{prox}_{\gamma i_S} = \begin{cases} a_i & \text{if} \quad x_i \leq a_i \\ x_i & \text{if} \quad x_i \in [a_i, b_i] \\ b_i & \text{if} \quad x_i \geq b_i \end{cases} \tag{16}$$

Derive expression for $\text{prox}_{\gamma i_S^*}$:

Similarly, from Fermat's rule and assuming $\partial i_S^* = \{a_i\}$ we can write equations 17 and 18.

$$0 = a_i + \gamma^{-1}(x_i - z_i) \iff -\gamma a_i = x_i - z_i \iff x_i = z_i - \gamma a_i \tag{17}$$

$$\Rightarrow x_i < 0 \text{ if } \gamma a_i > z_i \tag{18}$$

Instead assuming $\partial i_S^* = \{b_i\}$ and keeping Fermat's rule in mind we get equations 19 and 20.

$$0 = b_i + \gamma^{-1}(x_i - z_i) \iff x_i = z_i - \gamma b_i \tag{19}$$

$$\Rightarrow x_i > 0 \text{ if } \gamma b_i < z_i \tag{20}$$

Finally, assuming $x_i = 0 \Rightarrow \partial i_S^* = [a_i, b_i]$ yields equation 21.

$$0 = \nabla f(0) + \gamma^{-1}(0 - z) \Rightarrow z \in [\gamma a_i, \gamma b_i] \tag{21}$$

These equations implicates the full $\text{prox}_{\gamma i_S^*}$ as in equation 22.

$$\text{prox}_{\gamma i_S^*} = \begin{cases} z_i - \gamma a_i & \text{if} \quad z_i < \gamma a_i \\ 0 & \text{if} \quad z_i \in [\gamma a_i, \gamma b_i] \\ z_i - \gamma b_i & \text{if} \quad z_i > \gamma b_i \end{cases} \tag{22}$$

Task 4

Let y^* be a solution to the dual problem, derive an expression that gives a solution to the primal problem given y^*:

Dual problem is stated in equation 11. Assuming y^* is our solution, optimality conditions implicates equation 23.

$$\begin{cases} Lx \in \partial f^*(y^*) \\ x \in \partial g^*(-y^*) \end{cases} \tag{23}$$

We can see that the first row in equation 23 implicates 24, as $\partial f^*(y^*) = Q^{-1}(y^* - q)$ and L in an identity matrix.

$$x = Q^{-1}(y^* - q) \tag{24}$$

Task 5

See attached code.

Task 6

Trying a range of different step-sizes, resulted in Figures 1, 2, 3, 4, 5, 6, 7, 8 and 9. $\gamma = 2/L$ resulted in the fastest convergence whereas $\gamma = 2.5/L$ does not converge to zero, as the condition $0 \succ LI - Q$ does not hold. $\gamma = 2/L$ is not giving a solid theoretical convergence and may not converge to zero for all values, hence we are better of to choose $\gamma = 1.5/L$. A smaller step-size of $\gamma = 0.01/L$ converges to zero but more slowly.

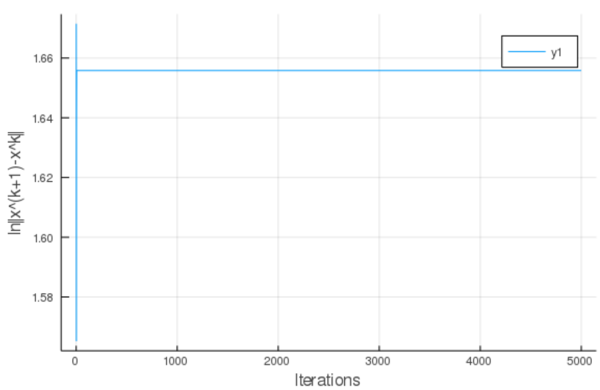

Figure 1: $\gamma = 10/L$

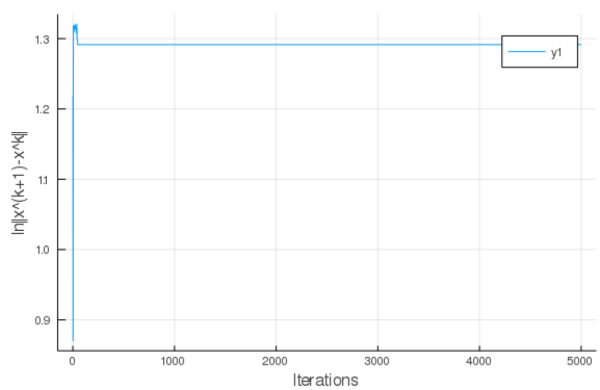

Figure 2: $\gamma = 3/L$

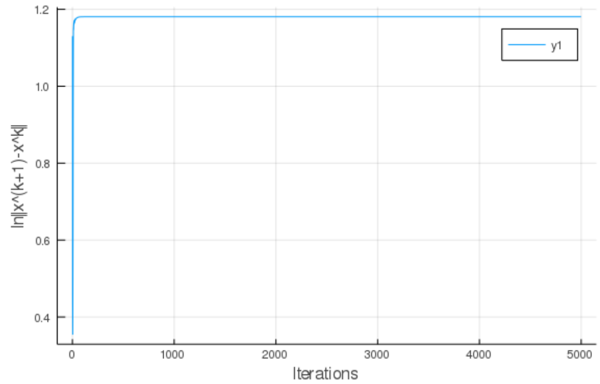

Figure 3: $\gamma = 2.5/L$

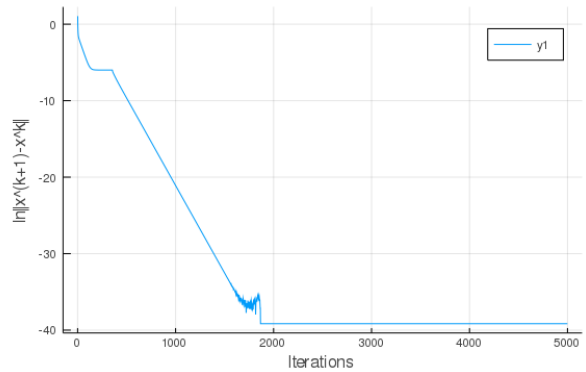

Figure 4: $\gamma = 2/L$

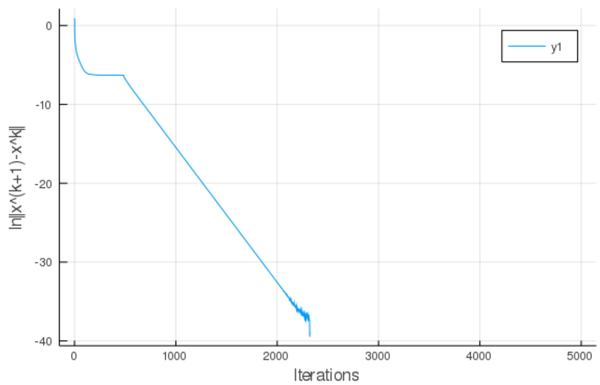

Figure 5: $\gamma = 1.5/L$

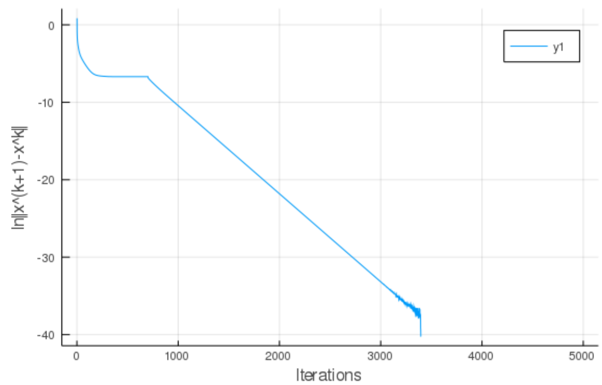

Figure 6: $\gamma = 1/L$

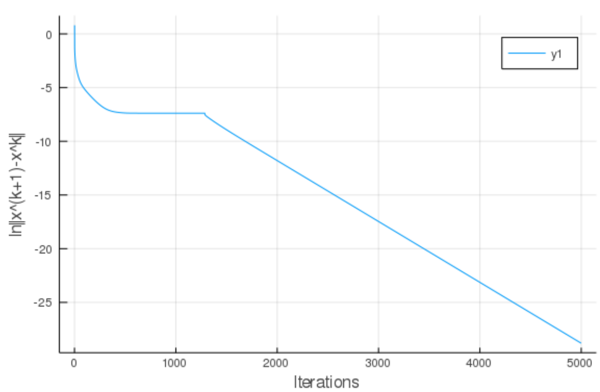

Figure 7: $\gamma = 0.5/L$

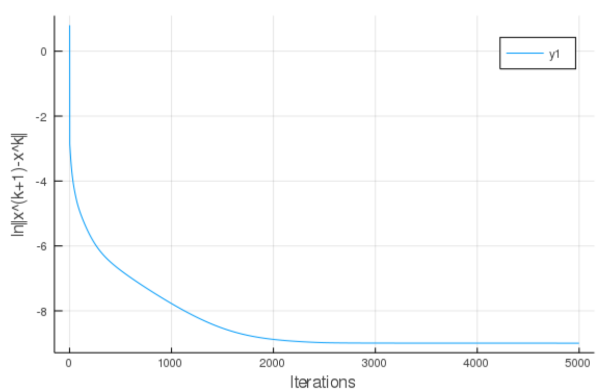

Figure 8: $\gamma = 0.1/L$

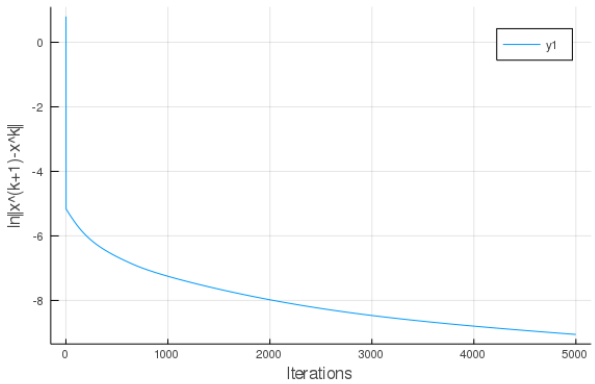

Figure 9: $\gamma = 0.01/L$

As shown in Figure 10, the first iteration will project it onto the set S. Since the prox operator for the indicator function is the projection onto the set we will never get an iterate outside the set. This can also be seen in Figure 11 where the optimization is executed with a to long step-size. The x-value does not converge, but it always remains in the set.

Since $\nabla^2 f = Q \succ 0$, the function is strongly convex. This means that there is one unique minimizer and this is what the proximal gradient method will find, independent on the initial value.

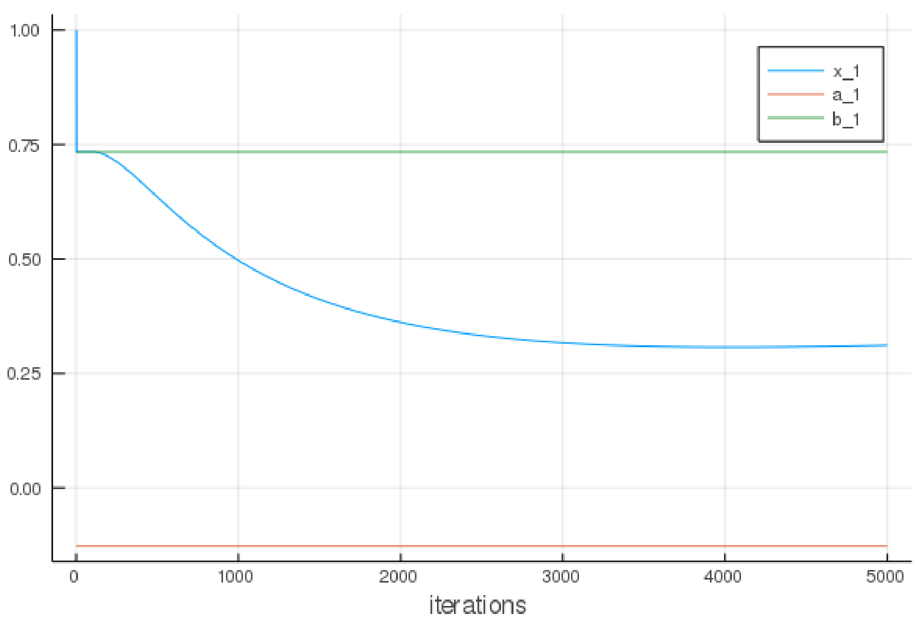

Figure 10: x_1 with boundary values with initial value x_0 outside S.

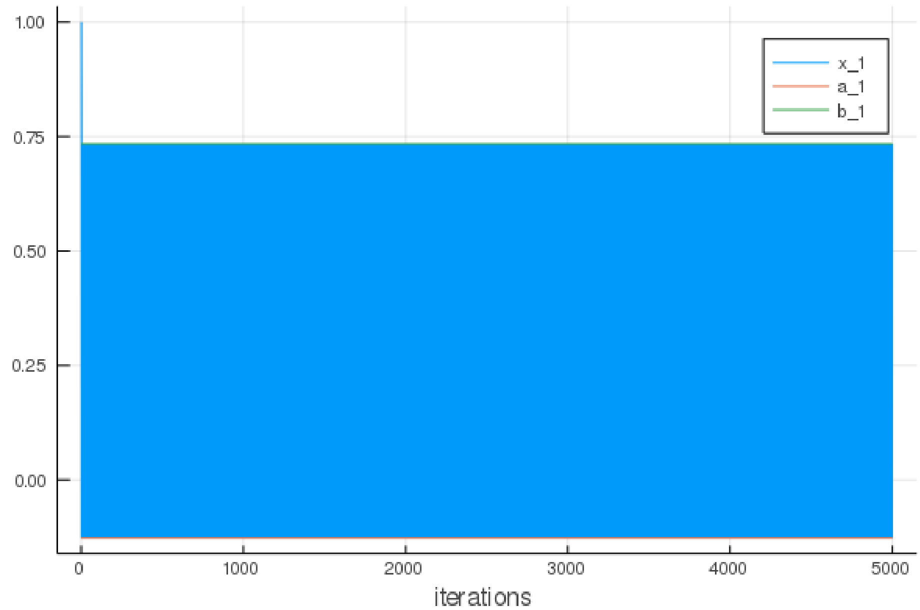

Figure 11: x_1 with boundary values and with high γ so problem does not converge.

Task 7

In Figures 12, 13, 14, 15, 16, 17, 18, 19 and 20 different step-sizes are evaluated for the dual problem. Figures 12, 13 and 14, it is clear that the iterations does not converge to zero. The convergence is faster for $\gamma = 2/L$ compared to lower gamma. Keep in mind Task 2 Equation 13 says that gamma might not converge at $2/L$ and a lower gamma should be picked, e.g. $\gamma = 1.5/L$

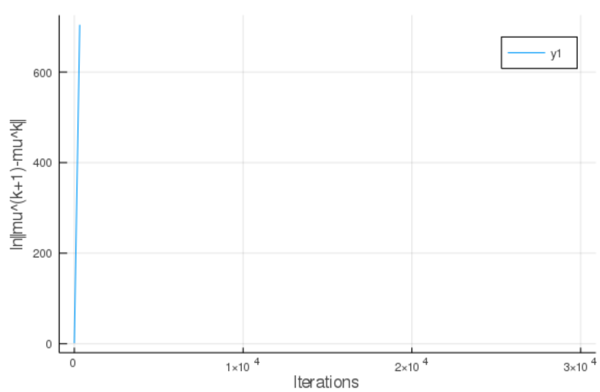

Figure 12: $\gamma = 10/L$

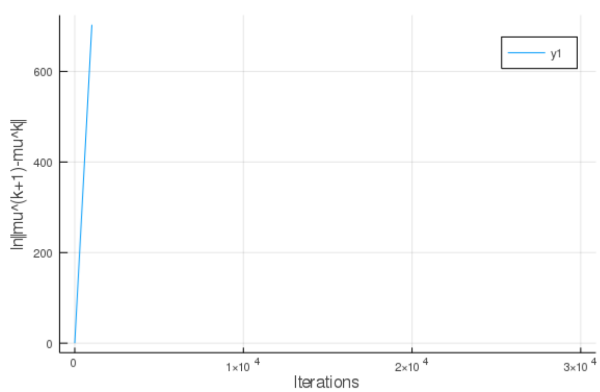

Figure 13: $\gamma = 3/L$

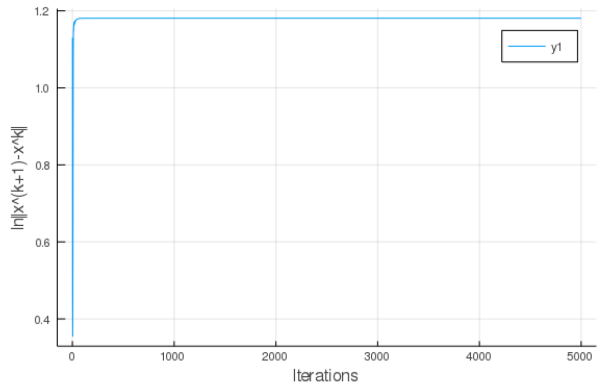

Figure 14: $\gamma = 2.5/L$

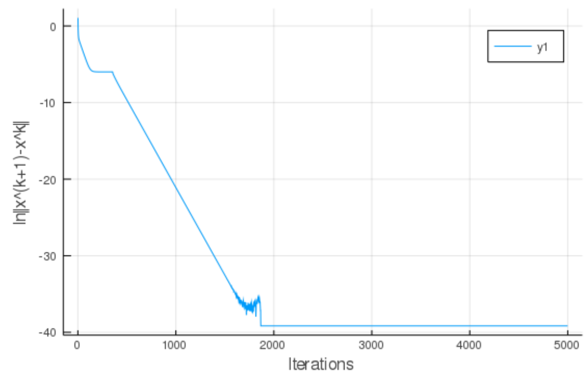

Figure 15: $\gamma = 2/L$

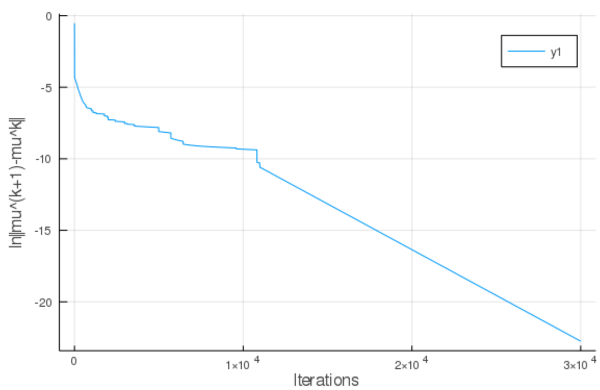

Figure 16: $\gamma = 1.5/L$

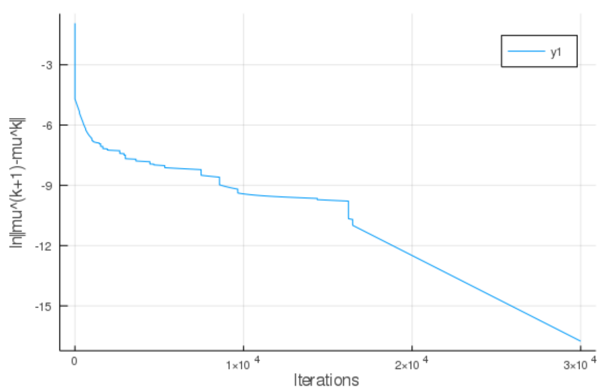

Figure 17: $\gamma = 1/L$

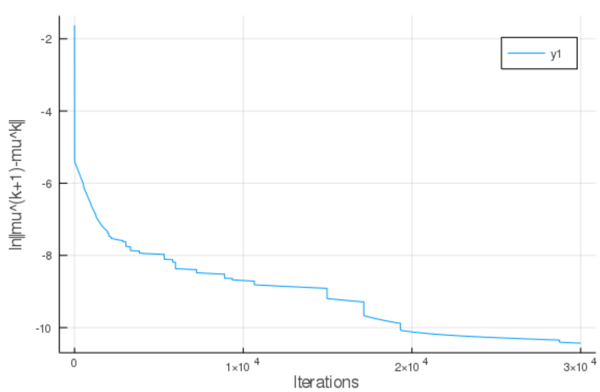

Figure 18: $\gamma = 0.5/L$

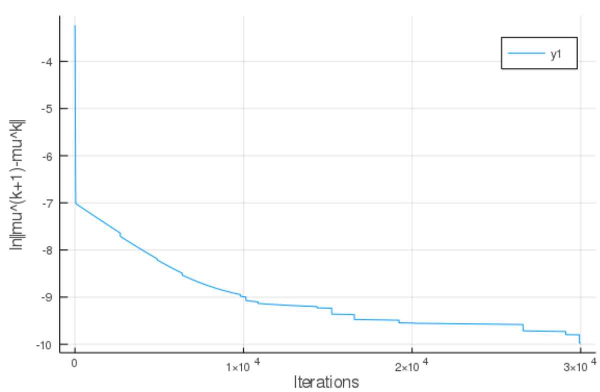

Figure 19: $\gamma = 0.1/L$

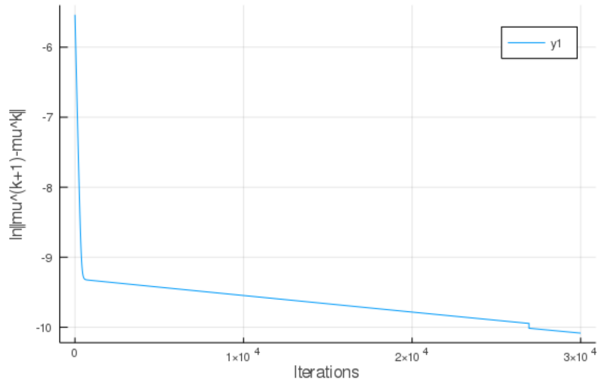

Figure 20: $\gamma = 0.01/L$

Our solution does not always satisfy the constraint $x \in S$, as seen when evaluating the indicator function. For instance the x_{17}-variable is slightly outside the set, i.e. $\hat{x}^k$ is not always converging. But as Figure 21 shown the distance from $\hat{x}^k$ and set S is converging towards zero for big k.

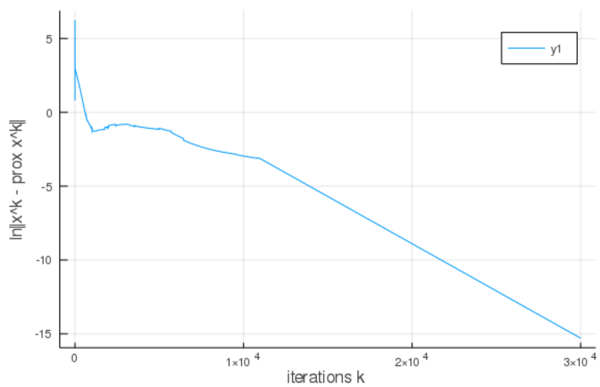

Figure 21: $f(\hat{x}^k)$

By evaluating the cost function in 1 it is clear that the solutions from the two approaches gives very similar solutions. The cost for the both the primal and the dual problem is -1.63505.

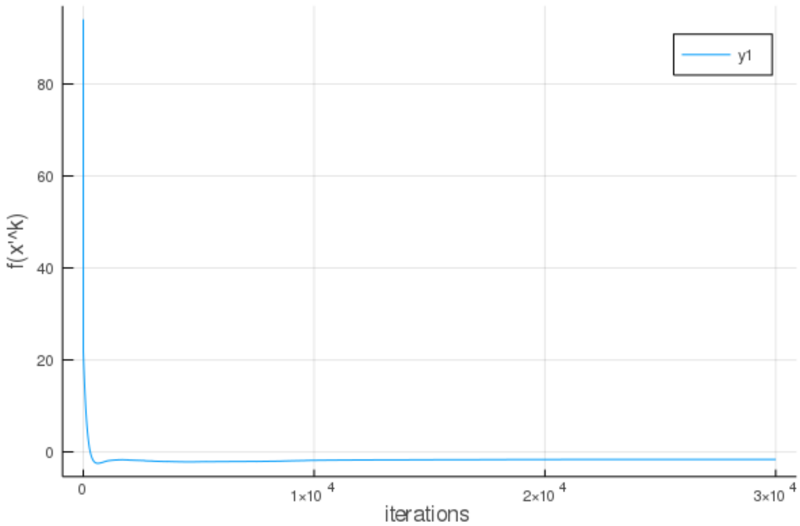

Figure 22: $f(\hat{x}^k)$