Subgradients Methods

Peter Lorenz

This documented is related to the programming part 4. The analytical part (point 1 and 2 on the assignment sheet) is in the other document. The point $x_k$is initialized by $(8, 8)^T$. The maximal number of iterations is limited to $10^5 \cdot 2.5$. A stoping criteria is set, if $f(x) - f^* < 10^{-3}$ then stop.

4. Plot the level sets of the function using matplotlib. You can use plt.contour() for this. Additionally, plot the iterates of the algorithm using the different step-sizes.

   (a) **Constant Stepsize**: $t_k = 10^{-3}$. The constant step size does not have enough epochs (only 10^6) available to converge (see Figure 1 and 2). The optimal point $x^* = (0.0004, .00113)^T$is reached.

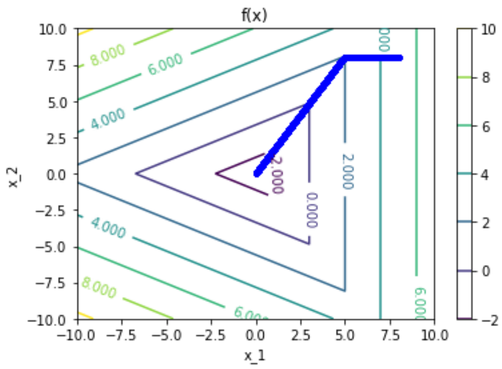

Figure 1: Contour plot.

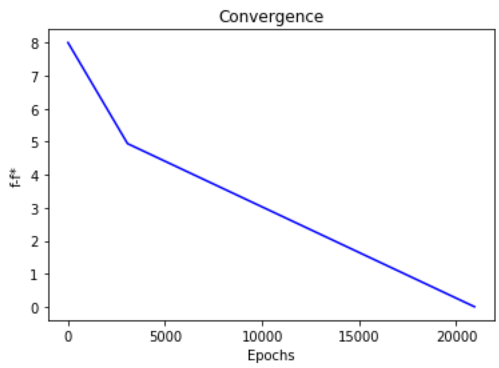

Figure 2: Convergence after >20.000 epochs.

   (b) **Diminishing Stepsize**: $t_k = \frac{\beta}{k+\gamma}$ with *k...current epoch*. A good parameter choice is $\beta = 5$ and $\gamma = \sqrt{\beta}$. The stopping criteria got active after ~550 epochs. Optimal point was found at $x^* = (0.0006, -0.0013)^T$.

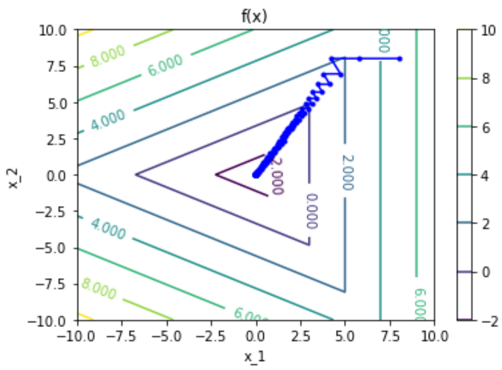

Figure 3: Contour plot.

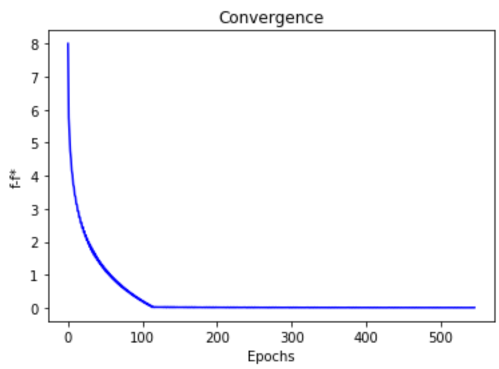

Figure 4: Convergence after 550 epochs.

   (c) **Dynamic Stepsize**: $t_k = \frac{f(x^k) - f^*}{||g^k||^2}$, with $f^* = f(x^*) = -3$ (see other document, analytical part). Additionaly I added a small ϵ to the denominator so that there is no division by 0 (as learned in the course Optimization for CS). The optimal point is found at $x^* = (1.06 \cdot 10^{-6}, 5.35 \cdot 10^{-4})^T$ by only 12 iterations.

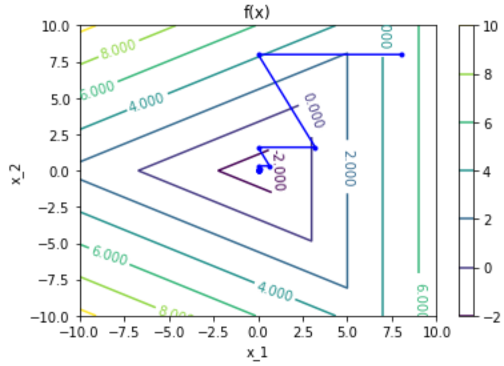

Figure 5: Contour plot.

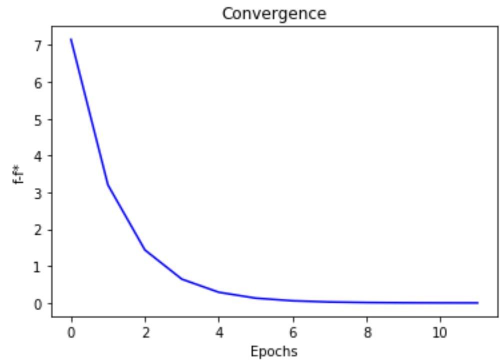

Figure 6: Convergence after 11 epochs.

5. Plot the convergence of $f(x) - f^*$. Which step size selection works best?

   The **Dynamic Stepsize** works best, because of the fast convergence rate and the ease of use (no paramaters have to be found as for Diminishing stepsize) if f^* is known.

6. Try to experiment with the dynamic step size when f^* is not known.

   Dynamic stepsize shows up good results, because of using the optimal point in the formula. This time, we do not know $f^*$and want to find out, if the diminishing stepsize still works.

   (a) Subtraction: If I change $f^* > 1$ only one epoch is done. So, I choose $f^* = 1$ for illustration. No good results can be seen, because it does not converge after 25.000 epochs.

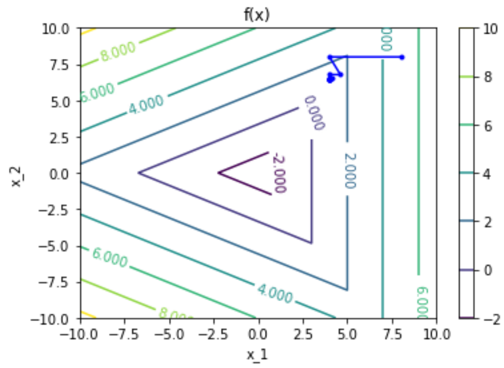

Figure 7: Contour plot, $f^* = 1$.

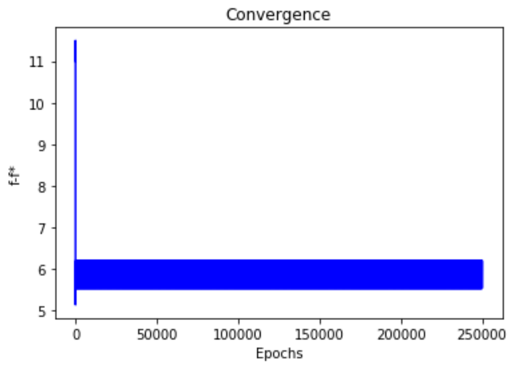

Figure 8: Convergence by $f^* = f(x^*) = 1$.

   (b) Addition: First, I subtract $f^* = -\frac{1}{2}$, the conpour plot Figure 9 looks simiar to Figure 7, but in Figure 10 can be seen that the algorithm converges faster through the stopping criteria. Nevertheless, still the optimum is not found and hence $f^* = -\frac{1}{2}$ or $f^* = -6$ are useless.

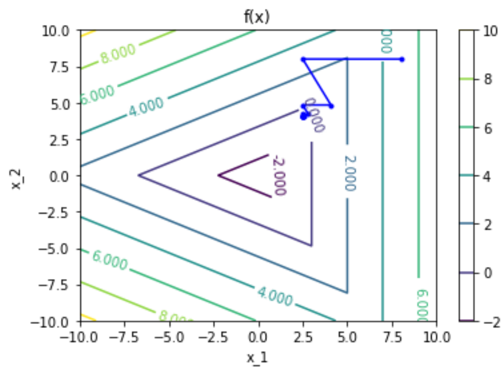

Figure 9: Contour plot, $f^* = -\frac{1}{2}$.

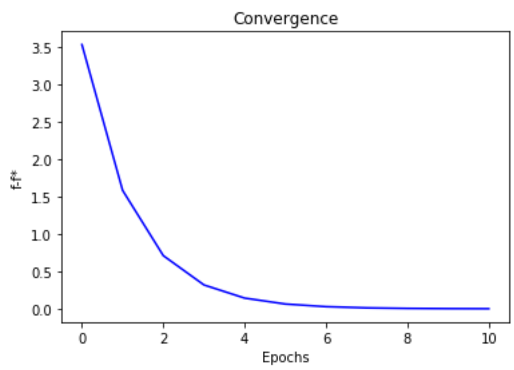

Figure 10: Convergence by $f^* = f(x^*) = -\frac{1}{2}$.

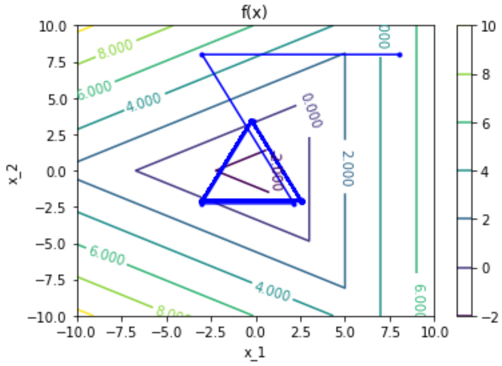

Figure 11: Contour plot, $f^* = -6$.

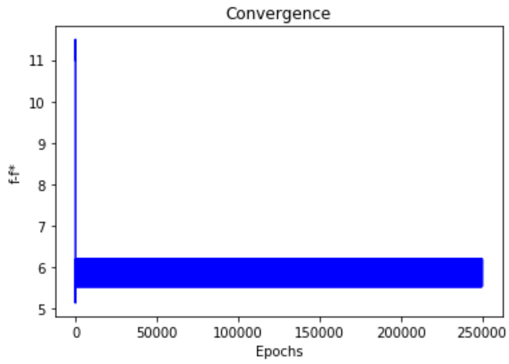

Figure 12: Convergence by $f^* = f(x^*) = -6$.

7. Experiment with different choices of the subgradient $g^k \in \partial f(x^k)$.

   The choice of different subgradients does not effect the result. I decided to take another 2 examples of subgradients for emphasizing my claim. Intitialization of x_k with different values does not change the result alot, because the algorithm converges with all chosen subgradients.

Choice	f1=f2	f1=f3	f2=f3	f1=f2=f3	Comment
1	$(0,0)^T$	$(0,0)^T$	$\left(-\frac{1}{\sqrt{5}}, 0\right)^T$	$(0,0)^T$	Used in previous numbers.
2	$\left(1, \frac{2}{\sqrt{5}}\right)^T$	$(1,0)^T$	$\left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)^T$	$\left(1, \frac{2}{\sqrt{5}}\right)^T$	Most right values in interval.
3	$\left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)^T$	$\left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)^T$	$\left(-\frac{1}{\sqrt{5}}, 0\right)^T$	$\left(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right)^T$	Most left values in interval.

Table 1: Choice of Subgradients

(a) Choice 2:

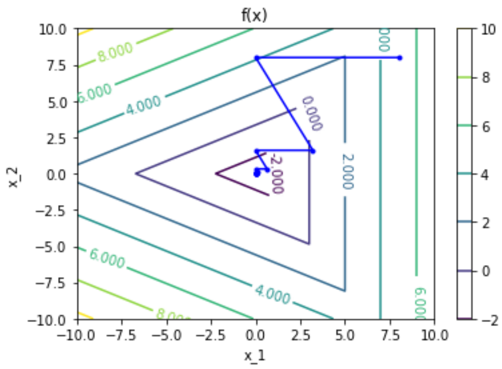

Figure 13: Contour plot.

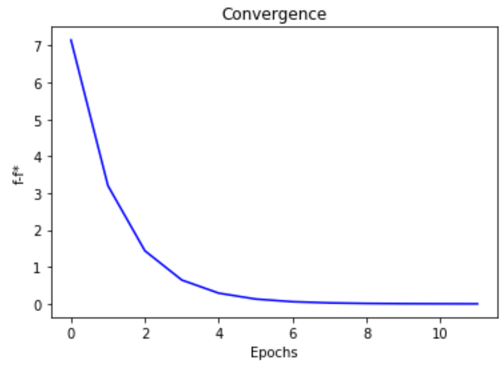

Figure 14: Convergence, $f^* = f(x^*) = -3$.

(b) Choice 3:

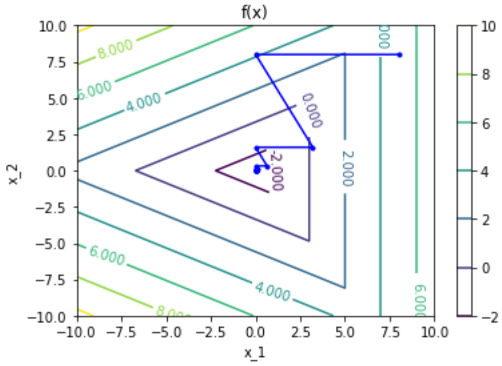

Figure 15: Contour plot.

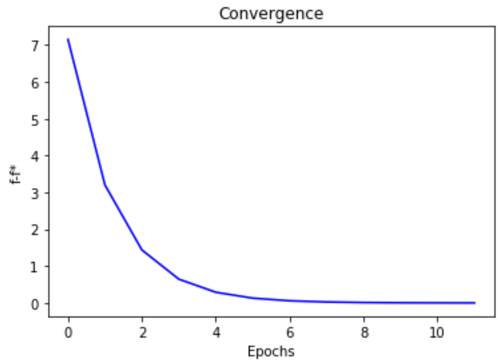

Figure 16: Convergence by $f^* = f(x^*) = -3$.