TMA4180 Optimization Theory
Spring 2013

Exercise set 3

Norwegian University of Science and Technology
Department of Mathematical Sciences

1 When using steepest descent, the choice of proper step sizes α_k is important. A temptingly simple approach is to simply decrease the step-size α_k until

$$f(x_k - \alpha_k \nabla f(x_k)) < f(x_k). \tag{1}$$

This however can fail already in simple examples. In this exercise, which is based on an example from [Ber], Section 1.2, we will investigate such a case.

Let

$$f(x) = \begin{cases} \dfrac{3(1-x)^2}{4} - 2(1-x) & \text{if } x > 1, \\ \dfrac{3(1+x)^2}{4} - 2(1+x) & \text{if } x < -1, \\ x^2 - 1. \end{cases}$$

a) Calculate the gradient of f.
Solution:

$$f'(x) = \begin{cases} -\dfrac{3}{2}(1-x) + 2 & \textit{if } x > 1, \\ \dfrac{3}{2}(1+x) - 2 & \textit{if } x < -1, \\ 2x. \end{cases}$$

b) Show that f is strictly convex.
Solution: Check the second derivative; it is everywhere positive.

c) Show that $f(x) < f(y)$ iff $|x| < |y|$.
Solution: f is an even function, so without loss of generality, let $0 < x < y$. Furthermore, it is strictly convex and on the positive x-axis strictly monotonically increasing.

d) Now prove that for $|x| > 1$: $f(x - \nabla f(x)) < f(x))$. Conclude that in general: $f(x_k - \nabla f(x_k)) < f(x_k)$ for $|x_0| > 1$ and $x_{k+1} := x_k - \nabla f(x_k)$.
Solution: From the previous problem, we know it suffices to show that $|x_k - \nabla f(x_k)| < |x_k|$. Without loss of generality, assume $x_k > 1$. We then get:

$$|x_k - \nabla f(x_k)| = |x_k + \frac{3}{2}(1-x) - 2| = |-\frac{1}{2}(x_k + 1)| < |x_k|.$$

Naively following the gradient flow without a careful linesearch will in this case lead us to oscillate between $x < -1$ and $x > 1$.

Now we want to solve the problem

$$\min_{x \in \mathbb{R}} f(x) \tag{2}$$

numerically using a gradient descent method.

e) Plot the function f.

f) Try to solve (2) using the steepest descent method. Implement a step-size reduction rule based on Inequality (1), i.e., decrease α_k by a given factor if (1) doesn't hold. Use a starting value $x_0 > 1$ and $\alpha_0 = 1$ as initial stepsize. Plot the first 10 iterations of the descent method. Explain the behaviour of the descent method for this function and the given starting value based on Problem d).

g) Now implement a step-size reduction method based on the Wolfe-Conditions (see N&W Equation (3.4) (p. 33 in 2nd edition)) and the backtracking algorithm 3.1 in N&W (p. 37 in 2nd edition). Use this to solve Problem (2). Use again a starting value $x_0 > 1$ and $\alpha_0 = 1$ as initial stepsize ($= \bar{\alpha}$ in algorithm 3.1; choose also suitable values for ρ and c). Plot again the iterates of the steppest descent method. Plot also the step-sizes as determined by the Wolfe-Condition and the backtracking algorithm.

References

[Ber] Bertsekas, Dimitri P., *Nonlinear Programming*, 2nd edition, 1999, Athena Scientific