Universität Konstanz Sommersemester 2014
Fachbereich Mathematik und Statistik
Prof. Dr. Stefan Volkwein
Roberta Mancini, Carmen Gräßle, Laura Lippmann, Kevin Sieg

Optimierung

http://www.math.uni-konstanz.de/numerik/personen/volkwein/teaching/

Sheet 6

Deadline for hand-in: 2014/07/16 at lecture

Exercise 18 (Scaled gradient method)
Consider the quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x,y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 100 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + 3,$$

and use a modified version of the Gradient Method where the update is

$$x^{k+1} = x^k - t^k M^{-1} \nabla f(x^k)$$

with t^k exact stepsize and M one of the following matrices

$$M = \mathrm{Id} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M = \nabla^2 f = \begin{pmatrix} 100 & -1 \\ -1 & 2 \end{pmatrix}, \quad M = \begin{pmatrix} f_{xx} & 0 \\ 0 & f_{yy} \end{pmatrix} = \begin{pmatrix} 100 & 0 \\ 0 & 2 \end{pmatrix}.$$

Use as basis the Gradient Method you implemented for the first program sheet to determine the number of gradient steps required for finding the minimum of f with the different matrices M and initial value `x0 = [1.5;0.6]`. Hand in suitable and informative plots and comment your observations (you don't need to hand in the code!).

Exercise 19 (Cauchy-step property) (4 Points)
The Cauchy step is defined as $s_a^c = -t_a \nabla f(x_a)$, where t_a is giving by (see the lecture notes)

$$t_a = \begin{cases} \frac{\Delta_a}{\|\nabla f(x_a)\|} & \text{if } \nabla f(x_a)^\top H_a \nabla f(x_a) \le 0, \\ \min\left(\frac{\Delta_a}{\|\nabla f(x_a)\|}, \frac{\|\nabla f(x_a)\|^2}{\nabla f(x_a)^\top H_a \nabla f(x_a)} \right) & \text{if } \nabla f(x_a)^\top H_a \nabla f(x_a) > 0. \end{cases}$$

Once the Cauchy point $x_a^c = x_a + s_a^c$ is computed, show that there is a sufficient decreasing in the quadratic model, i.e, the Cauchy step satisfies

$$\psi_a(0) - \psi_a(t_a) \ge \frac{1}{2} \|\nabla f(x_a)\| \min\left(\Delta_a, \frac{\|\nabla f(x_a)\|}{1 + \|H_a\|} \right).$$

Exercise 20 (Dogleg strategy)
Let us consider the quadratic model of the function f at iteration k in x_k

$$m_k(x) = f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{1}{2}(x - x_k)^\top B_k (x - x_k),$$

with B_k positive definite (Hessian matrix in x_k or an approximation of that). The trust-region subproblem gives back $x(\Delta)$ as

$$x(\Delta) = \underset{\|x-x_k\|<\Delta}{\arg\min}\ m_k(x).$$

If the trust region is big enough i.e, as if there is no constraint in the TR-subproblem, the solution $x(\Delta)$ would be the minimizer or $m_k(x)$: one would get the (quasi)Newton solution

$$x^{QN} = x_k - B_k^{-1}\nabla f(x_k).$$

When, instead, Δ is too small, the quadratic contribution is small and one tends to get a solution in the form given by the Cauchy point formula,

$$x^{CP} = x_k - \Delta\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|} \quad \text{or} \quad x^{CP} = x_k - \frac{\|\nabla f(x_k)\|^2}{\nabla f(x_k)^\top B_k \nabla f(x_k)}\nabla f(x_k).$$

Hence, for all the others Δ, $x(\Delta)$ will describe a curve in between the points x^{CP} and x^{QN}.

The idea of the dogleg method is to find an approximated solution replacing the curve just described with a path consisting of two straight lines: one from the current point x_k to the Cauchy-point x^{CP} and the other one from x^{CP} to the (quasi)Newton solution x^{QN}. The path is then described as

$$x(\tau) = \begin{cases} x_k + \tau(x^{CP} - x_k) & \tau \in [0,1] \\ x^{CP} + (\tau - 1)(x^{QN} - x^{CP}) & \tau \in [1,2]. \end{cases}$$

Show that $m(x(\tau))$ is a decreasing function of τ.

Hint: consider m on the two straight lines separately.