Prof. Dr. Karsten Urban
M. Sc. Lewin Ernst

Winter term 2024/2025
Sheet 05, 30 Points

Numerical Optimization
exercise sheet

review on 20.11.2024 during the exercise class

1. *(Convergence of nonlinear conjugate gradient method)*
In practice, the nonlinear cg-method is appealing for large nonlinear optimization problems, because each iteration requires only the evaluation of the objective function and its gradient. No matrix operations are required for the step computation, and just a few vectors have to be stored. One can prove a global convergence theorem (for the restarted version of the nonlinear cg-method, i.e. $\beta_{FR}^{(k)} \leftarrow 0$ for some k) based on Theorem 2.3.13 (Zoutendijk). For the Theorem to apply we need to choose decent directions $p^{(k)}$, i.e.

$$\nabla f(x^{(k)})^T p^{(k)} = -\|\nabla f(x^{(k)})\|^2 + \beta_{FR}^{(k)} \nabla f(x^{(k)})^T p^{(k-1)} < 0, \quad \forall k \in \mathbb{N}_0. \tag{1}$$

With the Lemma below and the strong Wolfe conditions with $0 < c_1 < c_2 < \frac{1}{2}$ one can prove that (1) is fulfilled and with the Theorem of Zoutendijk one conclude that the nonlinear cg-method is convergent.

Lemma 1. *Let the step sizes $\alpha^{(k)}$ of Algorithm 2.3.7 satisfy the strong Wolfe conditions with $0 < c_1 < c_2 < \frac{1}{2}$. Then the method generates directions $p^{(k)}$ that satisfy the following inequalities*

$$-\frac{1}{1-c_2} \leq \frac{\nabla f(x^{(k)})^T p^{(k)}}{\|\nabla f(x^{(k)})\|^2} \leq \frac{2c_2 - 1}{1 - c_2}, \quad \forall k \in \mathbb{N}_0$$

a) Proof Lemma 1 by induction.

b) Show that for $0 < c_2 < \frac{1}{2}$, we have

$$-1 < \frac{2c_2 - 1}{1 - c_2} < 0.$$

Remark: This means we have descent directions $p^{(k)}$.

c) Why does convergence by Zoutendijk not directly follow? Argue with $\cos\theta^{(k)}$.

(6 + 2 + 4 = 12 Points)

2. *(Convergence of trust-region methods)*
Trust-region methods are powerful global optimization schemes. The idea is in order to determine the next step $x^{(k+1)} = x^{(k)} + p^{(k)}$ we minimize the quadratic model

$$q^{(k)}(p) := f^{(k)} + (g^{(k)})^T p + \frac{1}{2} p^T H^{(k)} p,$$

where we define $f^{(k)} := f(x^{(k)})$, $g^{(k)} := \nabla f(x^{(k)})$ and $H^{(k)} := \nabla^2 f(x^{(k)})$, within a trust-region

$$B_{\Delta^{(k)}} := \{p \in \mathbb{R}^n : \|p\| \leq \Delta^{(k)}\}.$$

Therefore, we have to solve

$$\min_{p \in \mathbb{R}^n} q^{(k)}(p), \quad \text{s.t. } \|p\| \leq \Delta^{(k)}$$

in every iteration of the algorithm. Although, this problem can be solved efficiently the trust-region method is globally convergent when the approximate solution $p^{(k)}$ lies within the trust region and gives a *sufficient* reduction in the model. The *sufficient* reduction can be quantified in terms of the Cauchy point/Cauchy step (see Assumption 2.4.2). We can calculate the Cauchy point $p_C^{(k)}$ as follows. Determine the solution $p_S^{(k)}$ of the linear problem

$$\min_{p\in\mathbb{R}^n} f^{(k)} + (g^{(k)})^T p, \quad \text{s.t. } \|p\| \le \Delta^{(k)}$$

and then calculate the solution $\tau_C^{(k)} > 0$ of

$$\min_{\tau^{(k)}>0} q^{(k)}(\tau^{(k)} p_S^{(k)}), \quad \text{s.t. } \|\tau^{(k)} p_S^{(k)}\| \le \Delta^{(k)}.$$

a) Show that

$$p_S^{(k)} = -\frac{\Delta^{(k)}}{\|g^{(k)}\|} g^{(k)}$$

and

$$p_C^{(k)} = -\tau^{(k)} \frac{\Delta^{(k)}}{\|g^{(k)}\|} g^{(k)},$$

where

$$\tau^{(k)} = \begin{cases} 1, & \text{if } (g^{(k)})^T H^{(k)} g^{(k)} \le 0 \\ \min\{1, \|g^{(k)}\|^3 / \left(\Delta^{(k)} (g^{(k)})^T H^{(k)} g^{(k)}\right), & \text{otherwise} \end{cases}$$

Remark: This means that the computation of the Cauchy point is rather easy and we can check if Assumption 2.4.2 is fulfilled.

b) The proof of the global convergence theorem takes several auxiliary results. Therefore, we only prove the following: Let $f \in C^1(\mathbb{R}^n)$. If the trust-region algorithm produces a sequence $(\|g^{(k)}\|)_{k\in\mathbb{N}}$ with

$$\liminf_{k\to\infty} \|g^{(k)}\| = 0$$

and the iterates $x^{(k)}$ stay in a bounded set Ω, then there is *a* limit point $\bar{x}$ of $(x^{(k)})_{k\in\mathbb{N}}$ such that

$$g(\bar{x}) := \nabla f(\bar{x}) = 0.$$

Remark: Note that this is a weaker statement than Theorem 2.4.5 eq. (2.4.6).

(6 + 6 = 12 Points)

3. *(Nonlinear cg-method,* MATLAB*)*
Implement the nonlinear cg-method after Polak-Ribiére with a line search of your choice. Minimize the Rosenbrock function and compare the convergence behavior with one of the derivative-free algorithms and Newtons' method (with a line search algorithm of your choice). This should give you a feeling for the algorithms, which use no derivative information, only the gradient and also second order derivative information.

(6 Points)