Prof. Dr. Dr. h.c. Florian Rupp
Akaki Matcharashvili

General Descent Methods

This exercise sheet consists of two parts: at first problems for the Central Exercise class are given the solutions of which can serve you as further blueprints when solving similar tasks. These problems are discussed during the Central Exercise session. Then, the actual homework assignments are stated. Please, hand-in your results of the homework assignments through MSTeams at the date and time specified in MSTeams.

Central Exercise Problems:

Exercise 5.1: Criteria for Admissibility — Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, and (x^k), (s^k) and (σ_k) be generated by the general descent method. Prove the following assertions:

a) A sub-sequence of search directions $\{s^{k_j}\}$ is admissible, if there exists a strictly monotonically increasing function $\varphi : [0, \infty) \to [0, \infty)$ with $\varphi(0) = 0$, that satisfies

$$-\nabla f(x^{k_j})^T s^{k_j} \;\geq\; \varphi\Big(\|\nabla f(x^{k_j})\|\Big)\|s^{k_j}\| \quad \forall j\,.$$

b) A sub-sequence of step sizes (σ_{k_j}) is admissible, if for any $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$, such that the following implication holds:

$$\begin{aligned} &\frac{-\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|} \;\geq\; \varepsilon \text{ for infinitely many } j \\ \Longrightarrow\;\; & f(x^{k_j}) - f(x^{k_j} + \sigma_{k_j} s^{k_j}) \;\geq\; \delta(\varepsilon) \text{ for infinitely many } j. \end{aligned}$$

Exercise 5.2: Curry Step Size Rule — Let $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable, and $x^0 \in \mathbb{R}^n$ such that the levelset $\mathcal{N}_f(x^0) = \{x \in \mathbb{R}^n \,:\, f(x) \leq f(x^0)\}$ is compact. Let further $x \in \mathcal{N}_f(x^0)$ and $s \in \mathbb{R}^n$ a descent direction of f in x. The **Curry Step Size Rule** computes $\overline{\sigma} > 0$ as the smallest positive stationary point of the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by $\phi(\sigma) := f(x + \sigma s)$:

$$\overline{\sigma} \;:=\; \min\big\{\sigma > 0 \,:\, \phi'(\sigma) = 0\big\}\,.$$

a) Show well-posedness of the Curry step size rule.
Hint: Assume first, that there is no positive stationary point of ϕ, and use the mean value theorem to construct a contradiction to the compactness of $\mathcal{N}_f(x^0)$. Further show, that under all stationary points of ϕ there is actually a smallest one.

b) Show, using the Intermediate Value Theorem, that there is a smallest $0 < \hat{\sigma} < \overline{\sigma}$ such that $\nabla f(x + \hat{\sigma} s)^T s = \frac{1}{2}\nabla f(x)^T s$.

c) Show using part b): $f(x) - f(x + \overline{\sigma} s) \geq -\frac{1}{2}\hat{\sigma}\nabla f(x)^T s$.

d) Use compactness of $\mathcal{N}_f(x^0)$ and continuity of ∇f to show with help of part b), that for any $\varepsilon > 0$ there is a $\delta(\varepsilon) > 0$, independent of x and s, such that

$$-\tfrac{1}{2}\frac{\nabla f(x)^T s}{\|s\|} \;\geq\; \varepsilon \quad\Longrightarrow\quad \|\hat{\sigma} s\| \;\geq\; \delta(\varepsilon)\,.$$

e) Use parts c) and d), to show that any sub-sequence of step sizes $\{\sigma_k\}_K$, generated by the Gradient Descent Method with starting point x^0 and the Curry step size rule is admissible.

Homework Assignment:

Problem 5.1: Non-Admissible Descent Directions — We consider a general descent method for the function $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x) = \frac{1}{2}\|x\|^2$. We choose descent directions $s^k := g_\perp^k - \frac{1}{2^{k+3}}\nabla f(x^k)$. Here, we choose $g_\perp^k \in \mathbb{R}^2$ in such a way, that $g_\perp^k \perp \nabla f(x^k)$ and $\|s^k\| = \|\nabla f(x^k)\|$. Further, let $(\sigma_k)_{k\in\mathbb{N}}$ be a sequence of admissible step sizes and $x^0 \in \mathbb{R}^2 \setminus \{0\}$ an arbitrary initial point. We will show that the strict global minimum $\bar{x} = 0$ of f is not an accumulation point of the sequence of iterates $(x^k)_{k\in\mathbb{N}}$, even though all s^k are descent directions.

a) Show that the given directions s^k are indeed descent directions and that the sequence $(\|x^k\|)_{k\in\mathbb{N}_0}$ is monotonously decreasing.

b) Show that for $k \in \mathbb{N}_0$ the norms $\|x^{k+1}\|$ and $\|x^k\|$ satisfy the following relation:

$$\|x^{k+1}\|^2 \; = \; p(\sigma_k)\|x^k\|^2 \,,$$

where $p(\sigma_k)$ is some polynomial in the step size σ_k. From this relation deduce a (k-dependent) upper bound for the step size σ_k.

c) Show, that for all $k \in \mathbb{N}_0$ the estimate $\|x^{k+1}\| \geq \frac{1}{2}\|x^0\| > 0$ holds and thus no sub-sequence of $(x^k)_{k\in\mathbb{N}}$ can converge to the minimum $\overline{x} = 0$.

Problem 5.2: Non-Admissible Step Sizes for the Armijo Rule — Consider a general descent method for a continuously differentiable objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the descent directions $s_k = -2^{-k}\nabla f(x_k)$.

a) Assume that the algorithm does not terminate after a finite number of steps. Show, that under this assumption, any sub-sequence of $\{s^\ell\}_{\ell\in L}$ is admissible.

b) Show, that the Armijo rule for the above search directions, in general does not return admissible step sizes σ_k. To this end, like in the lecture, use the function $f(x) = \frac{x^2}{8}$ with a starting point $x_0 > 0$ and show that there is no admissible sub-sequence $\{s^\ell\}_{\ell\in L}$.
Hint: First show by induction that $0 < x^{k+1} < x^k$, and then by the definitions of the descent directions and $\sigma_k \in (0, 1]$ that it always holds $|x^0 - x^k| \leq \frac{x^0}{2} \Leftrightarrow x^k \geq \frac{x^0}{2}$.

c) What causes the non-admissibility of the step sizes?

d) Show that for any strictly monotonously increasing function $\varphi : [0, \infty) \to [0, \infty)$ the following condition is violated:

$$\|s^k\| \;\geq\; \varphi\left(\frac{-\nabla f(x^k)^T s^k}{\|s^k\|}\right) \qquad \text{for all } k \in \mathbb{N}_0\,.$$

Problem 5.3: Efficiency of Constant Step Size (old exam question, $\approx$ 25 min) — Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$-function with the property $\lim_{\|x\|\to\infty} f(x) = +\infty$.

a) Show that the level set $N_f(x) := \{y \in \mathbb{R}^n : f(y) \leq f(x)\}$ of f is compact for any $x \in \mathbb{R}^n$. What does this imply regarding the existence of minima?

b) Show, that for any $M > 0$ there exists a constant $L > 0$ with

$$f(y) \;\leq\; f(x) + \nabla f(x)^T(y - x) + \tfrac{1}{2}L\|x - y\|^2 \qquad \text{for all } x, y \in \mathbb{B}_M(0)\,,$$

where $\mathbb{B}_M(0)$ is a ball around 0 with radius M.
Hint: Note that the derivatives of f are bounded on compact sets.

c) Show using the inequality of b), that for all $R > 0$ there exists a constant $C > 0$ with

$$f(x) - f(x - t\nabla f(x)) \geq t\left(1 - \tfrac{1}{2}Ct\right)\|\nabla f(x)\|^2 \qquad \text{for all } x \in \mathbb{B}_R(0) \text{ and all } t \in [0, 1]\,.$$

d) Show using a) and c) that for every $x_0 \in \mathbb{R}^n$ there is an $\overline{\sigma} > 0$, such that the constant step size $\sigma_k = \sigma$ for all $\sigma \in (0, \overline{\sigma})$ is efficient for the gradient descent method in order to minimize f and starting at x_0.

e) Let $x_0 \in \mathbb{R}^n$ and $\sigma_k = \sigma \in (0, \overline{\sigma})$ as in d), and the gradient descent method minimizing f, starting at x_0 and using the step sizes σ_k does not terminate after a finite number of steps. Show using a theorem of the lecture, that all accumulation points of x_k are stationary.

Problem 5.4: (Convergence of an Unconstrained Optimization Algorithm (old exam question, $\approx$ 25 min) — Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x)\,.$$

Let $\{x^k\}_{k\in\mathbb{N}_0}$ be a sequence of iteration points generated by some algorithm for solving this problem, and suppose that it holds that $\{\nabla f(x^k)\}_{k\in\mathbb{N}_0} \to 0$, that is, the gradient value tends to zero (which of course is a favorable behavior of the algorithm). The question is what this means in terms of the convergence of the more important sequence $\{x^k\}_{k\in\mathbb{N}_0}$.

Consider therefore the sequence $\{x^k\}_{k\in\mathbb{N}_0}$, and also the sequence $\{f(x^k)\}_{k\in\mathbb{N}_0}$ of function values. Given the assumption that $\{\nabla f(x^k)\} \to 0$, is it true that $\{x^k\}_{k\in\mathbb{N}_0}$ and/ or $\{f(x^k)\}_{k\in\mathbb{N}_0}$ converges or are even bounded? Provide every possible case in terms of the convergence of these two sequences, and give examples, preferably simple ones for $n = 1$.