SoSe 2020 Kiel University
Dr. Joscha Reimer Department of Computer Science

Solution homework exercise 5.2:

Statement: Consider the general descent method applied to

$$f : \mathbb{R}^2 \to \mathbb{R}, x \mapsto \tfrac{1}{2}\|x\|_2^2$$

with $x_0 \in \mathbb{R}^2 \setminus \{0\}$ *and search direction*

$$d_k := s_k - \frac{1}{2^{k+3}} g_k$$

with $g_k := \nabla f(x_k)$ *and* $s_k \in \mathbb{R}^n$ *such that* $s_k \perp g_k$ *and* $\|d_k\|_2 = \|g_k\|_2$ *for all* $k \in \mathbb{N}_0$. *Let* $(x_k)_{k\in\mathbb{N}_0}$ *be the resulting sequence of iterates.*

a) $\rho_k < 2^{-(k+2)}$ *for all* $k \in \mathbb{N}_0$ *(where* ρ_k *are the step-sizes).*

b) $\|x_0 - x_{k+1}\|_2 \leq \frac{1}{2}\|x_0\|_2$ *for all* $k \in \mathbb{N}_0$.

c) $(x_k)_{k\in\mathbb{N}_0}$ *has no accumulation point which is a local minimum of* f.

d) d_k *are descent directions for all* $k \in \mathbb{N}_0$.

e) $(d_k)_{k\in\mathbb{N}_0}$ *is not gradient-related.*

Proof: Due to the definition of f and g_k

$$\nabla f(x) = x \text{ for all } x \in \mathbb{R}^2. \tag{1}$$

and thus

$$g_k = x_k \text{ for all } k \in \mathbb{N}_0. \tag{2}$$

Furthermore

$$\begin{aligned}
\nabla f(x_k)^T d_k &= g_k^T d_k && \text{(def. of } g_k) \\
&= g_k^T \left(s_k - \frac{1}{2^{k+3}} g_k\right) && \text{(def. of } d_k) \\
&= g_k^T s_k - \frac{1}{2^{k+3}} g_k^T g_k && (s_k \perp g_k) \\
&= -\frac{1}{2^{k+3}} \|g_k\|_2^2 && ((2)) \\
&= -\frac{1}{2^{k+3}} \|x_k\|_2^2.
\end{aligned} \tag{3}$$

a) Let $k \in \mathbb{N}_0$.

$$\begin{aligned}
\|g_k\|_2^2 &= \|x_k\|_2^2 && \text{((2))}\\
&= 2f(x_k) && \text{(def. of } f)\\
&> 2f(x_{k+1}) && \text{(def. of method)}\\
&= \|x_{k+1}\|_2^2 && \text{(def. of } f)\\
&= \|x_k + \rho_k d_k\|_2^2 && \text{(def. of method)}\\
&= \|g_k + \rho_k d_k\|_2^2 && \text{((2))}\\
&= \|g_k\|_2^2 + 2g_k^T \rho_k d_k + \rho_k^2 \|d_k\|_2^2 && \text{(calculation rule for norms)}\\
&= \|g_k\|_2^2 - 2\rho_k \frac{1}{2^{k+3}} \|g_k\|_2^2 + \rho_k^2 \|d_k\|_2^2 && \text{((3))}\\
&= \|g_k\|_2^2 - \rho_k \frac{1}{2^{k+2}} \|g_k\|_2^2 + \rho_k^2 \|g_k\|_2^2 && (\|d_k\|_2 = \|g_k\|_2)
\end{aligned}$$

This implies

$$\rho_k(2^{k-2} - \rho_k)\|g_k\|_2^2 > 0.$$

Thus

$$\rho_k < 2^{k-2}$$

because $\rho_k \geq 0$ and $\|g_k\|_2^2 \geq 0$.

b) Let $k \in \mathbb{N}_0$.

$$\begin{aligned}
\|x_0 - x_{k+1}\|_2 &= \left\| \sum_{i=1}^{k} (x_i - x_{i+1}) \right\|_2 && \text{(telescoping sum)}\\
&< \sum_{i=1}^{k} \|x_i - x_{i+1}\|_2 && \text{(triangle inequality)}\\
&= \sum_{i=1}^{k} \|\rho_i d_i\|_2 && \text{(def. of method)}\\
&= \sum_{i=1}^{k} \rho_i \|d_i\|_2 && \text{(norm property and } \rho_i \geq 0)\\
&= \sum_{i=1}^{k} \rho_i \|x_i\|_2 && (\|d_k\|_2 = \|g_k\|_2 \text{ and (2))}\\
&< \sum_{i=1}^{k} 2^{-(i+2)} \|x_i\|_2 && \text{(statement a))}\\
&< \sum_{i=1}^{k} 2^{-(i+2)} \|x_0\|_2 && (\|x_{i+1}\|_2 < \|x_i\|_2 \text{ since } f(x_{i+1}) < f(x_i))\\
&= \tfrac{1}{4}\|x_0\|_2 \sum_{i=1}^{k} \left(\tfrac{1}{2}\right)^i\\
&= \tfrac{1}{4}\|x_0\|_2 \frac{1 - \left(\frac{1}{2}\right)^{k+1}}{1 - \frac{1}{2}} && \text{(partial sum of geometric series)}\\
&= \tfrac{1}{2}\|x_0\|_2 \left(1 - \left(\tfrac{1}{2}\right)^{k+1}\right)\\
&< \tfrac{1}{2}\|x_0\|_2
\end{aligned}$$

c)
$$\nabla f(x) = x \text{ and } \nabla^2 f(x) = I \text{ for all } x \in \mathbb{R}^2, \tag{4}$$
implies that $x^* := 0$ is the only local minimum of f. Assume that x^* is an accumulation point of $(x_k)_{k\in\mathbb{N}_0}$. Then a subsequence $(x_{n_k})_{k\in\mathbb{N}_0}$ exists such that
$$\lim_{k\to\infty} x_{n_k} = x^*.$$
This implies
$$\lim_{k\to\infty} \|x_0 - x_{n_k}\|_2 = \|x_0\|_2.$$
But due to b) and $x_0 \neq 0$,
$$\|x_0 - x_{k+1}\|_2 \leq \tfrac{1}{2}\|x_0\|_2 < \|x_0\|_2 \text{ for all } k \in \mathbb{N}_0.$$
Thus is a contradiction and thus x^* can not be an accumulation point of $(x_k)_{k\in\mathbb{N}_0}$.

d) Let $k \in \mathbb{N}_0$. Due to c)
$$x_k \neq 0.$$
Thus (3) implies
$$\nabla f(x_k)^T d_k = -\frac{1}{2^{k+3}}\|x_k\|_2^2 < 0.$$

e) Due to $\|d_k\|_2 = \|g_k\|_2$ by definition and (3),
$$\|d_k\|_2 = \|g_k\|_2 = \|x_k\|_2$$
for all $k \in \mathbb{N}_0$. This implies with (4) and (3) that
$$\begin{aligned} -\frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|\|d_k\|} &= -\frac{-\frac{1}{2^{k+3}}\|x_k\|_2^2}{\|x_k\|\|x_k\|} \\ &= \frac{1}{2^{k+3}} \end{aligned}$$
for all $k \in \mathbb{N}_0$. Thus
$$\lim_{n\to\infty} -\frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|\|d_k\|} = 0.$$
Hence $(d_k)_{k\in\mathbb{N}_0}$ are not gradient-related.

□