E. Herberg, M. Marić, V. Stein
Heidelberg University

Exercise 4 - Solution

Date issued: 6th May 2024
Date due: 14th May 2024

Homework Problem 4.1 (Angle condition implies admissibility) 2 Points

Prove Lemma 4.4 from the lecture notes, i.e. if the angle condition (4.8) holds with some $\eta \in (0, 1)$, then the sequence $d^{(k)}$ of search directions is admissible.

Solution.

We have

$$f'(x^{(k)})\, d^{(k)} = \big(\nabla f(x^{(k)}), d^{(k)}\big) = \big(\nabla_M f(x^{(k)}), d^{(k)}\big)_M.$$

The angle condition (4.8) implies

$$-\frac{f'(x^{(k)})\, d^{(k)}}{\|d^{(k)}\|_M} \geq \eta\, \|\nabla_M f(x^{(k)})\|_M = \eta\, \|f'(x^{(k)})^\intercal\|_{M^{-1}} \geq 0.$$

(1 Point)

When the left-hand term goes to zero, then $f'(x^{(k)})$ must go to zero as well. (1 Point)

Homework Problem 4.2 (Efficient Step Sizes in Quadratic Steepest Descent) 4 Points

Show that both constant step sizes (as in § 3.3) as well as the Cauchy step size are efficient for the steepest descent method (cf. Algorithm 3.6) for solving quadratic optimization problems of the type

$$\text{minimize} \quad f(x) := \frac{1}{2} x^\intercal A x - b^\intercal x + c \quad \text{where } x \in \mathbb{R}^n$$

with s. p. d. $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

Solution.

For efficiency of step sizes α, we need to show that there exists a $\theta > 0$ such that, as long as $d^{(k)} \neq 0$, we have that

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) - \theta \left(\frac{f'(x^{(k)})\, d^{(k)}}{\|d^{(k)}\|_M} \right)^2 \tag{4.11}$$

for all $k \geq 0$.

In the steepest descent scheme for quadratic functionals, we use

$$d^{(k)} = -M^{-1} r^{(k)} = -M^{-1}(Ax^{(k)} - b) = -M^{-1} f'(x^{(k)})^\intercal$$

and therefore obtain

$$\begin{aligned} f(x^{(k+1)}) - f(x^{(k)}) &= f(x^{(k)} + \alpha^{(k)} d^{(k)}) - f(x^{(k)}) \\ &= \frac{1}{2} d^{(k)\intercal} A d^{(k)}\, \alpha^{(k)2} + (Ax^{(k)} - b)^\intercal d^{(k)} \alpha^{(k)} \\ &= \frac{1}{2} \|d^{(k)}\|_A^2 \alpha^{(k)2} - \|d^{(k)}\|_M^2 \alpha^{(k)}. \end{aligned}$$

(2 Points)

The Cauchy step size corresponding to $d^{(k)}$ are

$$\alpha^{(k)} \left(d^{(k)} \right) = -\frac{f'(x^{(k)}) d^{(k)}}{d^{(k)\intercal} A d^{(k)}} = \frac{\|d^{(k)}\|_M^2}{\|d^{(k)}\|_A^2}$$

(**Note:** This shows how the Cauchy step sizes measure the distortion of the M vs. the A isolines) Accordingly:

$$\begin{aligned} f(x^{(k+1)}) - f(x^{(k)}) &= \frac{1}{2} \|d^{(k)}\|_A^2 \alpha^{(k)2} - \|d^{(k)}\|_M^2 \alpha^{(k)} \\ &= -\frac{1}{2} \frac{\|d^{(k)}\|_M^4}{\|d^{(k)}\|_A^2} \\ &= -\frac{1}{2} \left(\frac{f'(x^{(k)}) d^{(k)}}{\|d^{(k)}\|_A} \right)^2 \\ &\leq -\frac{1}{2} \frac{1}{\lambda_{\max}(A; M)} \left(\frac{f'(x^{(k)}) d^{(k)}}{\|d^{(k)}\|_M} \right)^2 \end{aligned}$$

shows that the Cauchy step sizes are efficient for the M-steepest descent directions with $\theta = \frac{1}{2\lambda_{\max}(A;M)^2}$. (1 Point)

For $\alpha^{(k)} = \alpha$ constant, we obtain

$$
\begin{aligned}
f(x^{(k+1)}) - f(x^{(k)}) &= \frac{1}{2} \|d^{(k)}\|_A{}^2 \alpha^2 - \|d^{(k)}\|_M^2 \alpha \\
&\le \alpha \left(\frac{\alpha \lambda_{\max}(A;M)}{2} - 1 \right) \|d^{(k)}\|_M^2 \\
&= \alpha \left(\frac{\alpha \lambda_{\max}(A;M)}{2} - 1 \right) \left(\frac{f'(x^{(k)}) d^{(k)}}{\|d^{(k)}\|_M} \right)^2
\end{aligned}
$$

where

$$
\theta := -\alpha^{(k)} \left(\frac{\alpha^{(k)} \lambda_{\max}(A;M)}{2} - 1 \right)
$$

is positive whenever $\alpha \in (0, \frac{2}{\lambda_{\max}(A;M)})$, which is exactly the condition we derived for the convergence of the steepest descent scheme with constant step sizes in § 3.3. (1 Point)

Homework Problem 4.3 (Efficiency implies admissibility) 2 Points

Prove Lemma 4.8, i.e. if the sequence of step sizes $\alpha^{(k)}$ is efficient, then it is also admissible.

Solution.

Suppose that $\alpha^{(k)}$ is efficient, i. e.,

$$
0 \le \theta \left(\frac{f'(x^{(k)})\, d^{(k)}}{\|d^{(k)}\|_M} \right)^2 \le f(x^{(k)}) - f(x^{(k)} + \alpha^{(k)} d^{(k)})
$$

Therefore (4.10a) is clear. (1 Point)

To show (4.10b), suppose

$$
f(x^{(k)} + \alpha^{(k)} d^{(k)}) - f(x^{(k)}) \to 0.
$$

Since θ is strictly positive, this implies

$$
\frac{f'(x^{(k)})\, d^{(k)}}{\|d^{(k)}\|_M} \to 0,
$$

which confirms (4.10b). (1 Point)

Homework Problem 4.4 (Armijo steplength is not admissible in general) 6 Points

Show that the Armijo step sizes are not admissible in general.
Consider the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) := \frac{x^2}{8}$ with $M = 1$, admissible search directions $d^{(k)} = -2^{-k} f'(x^{(k)})$ and initial trial step size $\alpha^{(k,0)} = 1$ in the Armijo backtracking line search Algorithm 4.11.

Hint: Show that for any starting point $x^{(0)} > 0$ the generated sequence $(x^{(k)})$ by Algorithm 4.2 is monotonically decreasing, but converges to $\bar{x} \geq \frac{x^{(0)}}{2}$. Then explain, why $\bar{x}$ can not be a stationary point and how this shows that the Armijo step sizes are not admissible.

Solution.

We have $f'(x) = \frac{x}{4}$ and therefore $d^{(k)} = -2^{-k}\frac{x^{(k)}}{4} = -2^{-k-2}x^{(k)}$.
First, we show by induction that $(x^{(k)})$ is monotonically decreasing, i.e. $0 < x^{(k+1)} < x^{(k)}$ for all k.
For $k = 0$ it holds $x^{(0)} > 0$ by choice and we have due to $\alpha^{(0)} \in (0, 1]$:

$$x^{(1)} = x^{(0)} + \alpha^{(0)} d^{(0)} = x^{(0)} - 2^{-2}\alpha^{(0)} x^{(0)} = x^{(0)} \underbrace{\left(1 - \frac{\alpha^{(0)}}{4}\right)}_{<1} < x^{(0)}.$$

Now let $0 < x^{(k+1)} < x^{(k)}$ hold for some k and we show the claim for $k + 1$:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)} = x^{(k)} \underbrace{\left(1 - \frac{\alpha^{(k)}}{2^{k+2}}\right)}_{<1} < x^{(k)}$$

and also by assumption and $\alpha^{(k)} \in (0, 1]$

$$x^{(k+1)} = \underbrace{x^{(k)}}_{>0} \underbrace{\left(1 - \frac{\alpha^{(k)}}{2^{k+2}}\right)}_{>0} > 0.$$

This shows the claim. (2 Points)

Next, we show that $x^{(k)} > \frac{x^{(0)}}{2}$ for all k:

$$\begin{aligned} x^{(k)} &= \left(1 - \frac{\alpha^{(k-1)}}{2^{k-1+2}}\right) x^{(k-1)} \\ &= \prod_{i=0}^{k-1} \left(1 - \frac{\alpha^{(i)}}{2^{i+2}}\right) x^{(0)} \\ &\geq \prod_{i=0}^{k-1} \left(1 - \frac{1}{2^{i+2}}\right) x^{(0)} \\ &\geq \prod_{i=0}^{\infty} \left(1 - \frac{1}{2^{i+2}}\right) x^{(0)} \\ &\approx 0.5775\, x^{(0)} \\ &> \frac{x^{(0)}}{2}. \end{aligned}$$

We used that $\alpha^{(i)} \in (0,1]$ and $\left(1 - \frac{1}{2^{i+2}}\right) \in (0,1)$ for any i. (1 Point)

Let $\bar{x} = \lim_{k\to\infty} x^{(k)}$, then $0 < \frac{x^{(0)}}{2} \leq \bar{x} \leq x^{(0)}$. The only stationary point of f is given by $f'(x^*) = \frac{x^*}{4} = 0$, i.e. $x^* = 0$, hence $\bar{x}$ can not be a stationary point. (1 Point)

Also, we know that $f(x) < f(y)$ for $0 < x < y$ and $f(x) = 0$ if and only if $x = 0$. Hence, we have by continuity of f

$$0 < f\left(\frac{x^{(0)}}{2}\right) \leq f(x^{(k)}) \leq f(x^{(0)}) \qquad \forall k.$$

From $f(x^{(k)}) > 0 > -\infty$ for all k we deduce

$$\begin{aligned} & \underbrace{\lim_{k\to\infty} f(x^{(k+1)}) - f(x^{(0)})}_{=} > 0 - f(x^{(0)}) > -\infty \\ \overset{x^{(k+1)}<x^{(k)}}{\Rightarrow} \quad & 0 \geq \sum_{i=0}^{\infty} \left(f(x^{(i+1)}) - f(x^{(i)})\right) > -\infty \\ \Rightarrow \quad & \lim_{k\to\infty} f(x^{(k+1)}) - f(x^{(k)}) = 0 \end{aligned}$$

(1 Point)

If the step sizes were admissible, this would directly induce

$$\frac{f'(x^{(k)}) d^{(k)}}{\|d^{(k)}\|_M} \xrightarrow{k\to\infty} 0.$$

However, it holds

$$\frac{f'(x^{(k)})d^{(k)}}{\|d^{(k)}\|_M} = \frac{\frac{x^{(k)}}{4}(-2^{-k-2}x^{(k)})}{\|2^{-k-2}x^{(k)}\|_M} = \frac{-2^{-k-4}\|x^{(k)}\|_M^2}{2^{-k-2}\|x^{(k)}\|_M} = -\frac{1}{4}\|x^{(k)}\|_M < 0.$$

This shows that the step sizes can not be admissible. (1 Point)

Please submit your solutions as a single pdf and an archive of programs via moodle.