V. Schulz
M. Klar

SS 2022
Submission: 10.05.21, until 12:15

Numerical Optimization - Sheet 5

If you are a student in mathematics please solve the exercises with no tag and the ones with the tag **Mathematics**. If you are a data science student please solve the problems with no tag and those with the tag **Data Science**. Submissions with tags other than your subject count as bonus points. The tag **Programming** marks programming exercises.

Ex 1 **Data Science** (4 Points + *2 Bonus Points)

Let the norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ be equivalent as defined in the lecture. Show the following properties for a sequence of vectors $\{x^k\}_{k=1}^\infty$ converging to $\hat{x}$:

(i) Superlinear convergence in norm $\|\cdot\|_\alpha$ implies superlinear convergence in norm $\|\cdot\|_\beta$.

(ii) Sublinear convergence in norm $\|\cdot\|_\alpha$ implies sublinear convergence in norm $\|\cdot\|_\beta$.

(iii) ***** What (highly restrictive) additional condition do we need to ensure that the same implication holds for linear convergence?

Solution 1:
Equivalence of norms means that there is an $M > 0$ such that

$$\frac{1}{M}\|\cdot\|_\alpha \leq \|\cdot\|_\beta \leq M\|\cdot\|_\alpha$$
$$\left(\Leftrightarrow \frac{1}{M}\|\cdot\|_\beta \leq \|\cdot\|_\alpha \leq M\|\cdot\|_\beta\right)$$

(i) Superlinear $\|x^{k+1} - \hat{x}\| \leq \alpha_k\|x^k - \hat{x}\|$, for $\alpha_k \downarrow 0$:
$\|x^{k+1} - \hat{x}\|_\beta \leq M\|x^{k+1} - \hat{x}\|_\alpha \leq \alpha_k M\|x^k - \hat{x}\|_\alpha \leq \alpha_k M^2\|x^k - \hat{x}\|_\beta$. Which implies superlinear convergence as $M^2\alpha_k \downarrow 0$.

(ii) Sublinear $\|x^{k+1} - \hat{x}\| \leq \frac{C}{k^s}$:
$\|x^{k+1} - \hat{x}\|_\beta \leq M\|x^{k+1} - \hat{x}\|_\alpha \leq \frac{MC}{k^s}$

(iii) Linear $\|x^{k+1} - \hat{x}\| \leq \alpha\|x^k - \hat{x}\|$ for $\alpha < 1$:
$\|x^{k+1} - \hat{x}\|_\beta \leq M\|x^{k+1} - \hat{x}\|_\alpha \leq M\alpha\|x^k - \hat{x}\|_\alpha \leq \alpha M^2\|x^k - \hat{x}\|_\beta$
Hence, we need, that $\alpha M^2 < 1$, such that linear convergence can be implied.

Ex 2 **Mathematics** (4 Points)

Assume we apply the gradient descent method with exact line search to the problem

$$\min\ f(x) = \frac{1}{2}x^\top Qx$$

where $f : \mathbb{R}^n \to \mathbb{R}$, the scalar product is $(\cdot,\cdot) = (\cdot,\cdot)_2$, and $Q \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix. Please show that

$$f\left(x^{k+1}\right) = \left(1 - \frac{\left(g_k^\top g_k\right)^2}{\left(g_k^\top Qg_k\right)\left(g_k^\top Q^{-1}g_k\right)}\right) f\left(x^k\right)$$

where $g_k = \nabla f(x^k) = Qx^k$.

Solution 2:

Exact line search steepest descent: $t_k = \arg\min_t f(x_k - tg_k)$.

$$\begin{aligned}
t_k &= \arg\min_t \left(\frac{1}{2}(x_k - tg_k)^\top Q(x_k - tg_k)\right) \\
&= \arg\min_t \underbrace{\left(\frac{1}{2}x_k^\top Qx_k - tg_k^\top Qx_k + \frac{1}{2}t^2 g_k^\top Qg_k\right)}_{=:T(t)}
\end{aligned}$$

We obtain, that the step length α is

$$\begin{aligned}
&\frac{d}{dt}T(t_k) \stackrel{!}{=} 0 \\
&\Leftrightarrow -g_k^\top g_k + t_k g_k^\top Qg_k = 0 \\
&\Leftrightarrow t_k = \frac{g_k^\top g_k}{g_k^\top Qg_k} =: \alpha.
\end{aligned}$$

Hence, the gradient step of the exact line search is given by

$$\boxed{x_{k+1} = x_k - \alpha g_k = x_k - \frac{g_k^\top g_k}{g_k^\top Qg_k} g_k}.$$

If we apply f to x_{k+1} we obtain

$$\begin{aligned}
f(x_{k+1}) = f(x_k - \alpha g_k) &= \frac{1}{2}(x_k - \alpha g_k)^\top Q(x_k - \alpha g_k) \\
&= \frac{1}{2}x_k^\top Qx_k - \alpha x_k^\top Q g_k + \frac{1}{2}\alpha^2 g_k^\top Qg_k \\
&= \frac{1}{2}x_k^\top Qx_k - \alpha g_k^\top g_k + \frac{1}{2}\alpha^2 g_k^\top Qg_k \\
&= f(x_k) - \frac{(g_k^\top g_k)^2}{g_k^\top Qg_k} + \frac{1}{2}\frac{(g_k^\top g_k)^2}{(g_k^\top Qg_k)^2}(g_k^\top Qg_k) \\
&= f(x_k) - \frac{(g_k^\top g_k)^2}{g_k^\top Qg_k} + \frac{1}{2}\frac{(g_k^\top g_k)^2}{(g_k^\top Qg_k)} \\
&= f(x_k) - \frac{1}{2}\frac{(g_k^\top g_k)^2}{g_k^\top Qg_k} \\
&= f(x_k) - \frac{1}{2}\frac{(g_k^\top g_k)^2}{g_k^\top Qg_k} \quad \frac{g_k^\top Q^{-1}g_k}{g_k^\top Q^{-1}g_k} \\
&= f(x_k) - \frac{(g_k^\top g_k)^2}{(g_k^\top Qg_k)(g_k^\top Q^{-1}g_k)}\frac{1}{2}(g_k{}^\top Q^{-1}g_k) \\
&= f(x_k) - \frac{(g_k^\top g_k)^2}{(g_k^\top Qg_k)(g_k^\top Q^{-1}g_k)}\frac{1}{2}(x_k^\top Q Q^{-1}Qx_k) \\
&= f(x_k) - \frac{(g_k^\top g_k)^2}{(g_k^\top Qg_k)(g_k^\top Q^{-1}g_k)}\frac{1}{2}(x_k^\top Qx_k) \\
&= f(x_k) - \frac{(g_k^\top g_k)^2}{(g_k^\top Qg_k)(g_k^\top Q^{-1}g_k)} f(x_k) \Rightarrow \text{ Beh.}
\end{aligned}$$

Ex 3 (*4 Bonus Points)

The introduction of the Wolfe conditions raises the question whether we can guarantee that it is possible to find a point which fulfills them.
Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and d a descent direction at point x, i.e. $\nabla f(x)^T d < 0$. Let f additionally be lower bounded on the ray $\{x + \alpha d \,|\, \alpha > 0\}$. Show in the following steps that for a fixed $0 < c_1 < c_2 < 1$ there is a step size $\alpha > 0$ such that the Wolfe conditions

$$\phi(\alpha) := f(x + \alpha d) \le f(x) + c_1 \alpha \nabla f(x)^T d =: \ell(\alpha), \tag{1}$$
$$\phi'(\alpha) = \nabla f(x + \alpha d)^T d \ge c_2 \nabla f(x)^T d, \tag{2}$$

are fulfilled.

(i) Show that there is a smallest step size $\beta > 0$ such that

$$\phi(\beta) = \ell(\beta),$$

and that the inequality

$$\phi(\beta') < \ell(\beta')$$

holds for all $\beta' \in (0, \beta)$.

(ii) Deduce that there is an $\alpha \in (0, \beta)$ which fulfills

$$\phi'(\alpha) = \ell'(\alpha).$$

(iii) Conclude the assertion from (i) and (ii).

Solution 3:

(i) Denote $h(\beta) := \phi(\beta) - \ell(\beta)$.
a) Thus $h(0) = 0$ and

$$h'(0) = \phi'(0) - \ell'(0) = (1 - c_1)\nabla f(x)^T d < 0,$$

since d is an direction of descent and $c_1 < 1$.
$\Rightarrow \exists_{\beta_L > 0} \forall_{\beta' \in (0, \beta_L)} : h(\beta') < 0$ (β_L forms a lower bound for the intersection).

b) Since ϕ is bounded from below and continuous and ℓ is monotonically decreasing and linear, there exists an intersection $\beta_R \ge \beta_L$ with

$$h(\beta_R) = 0,$$

which forms an upper bound for the smallest intersection point.
c) Since $[\beta_L, \infty) \cap \{\beta' \,|\, h(\beta') = 0\}$ is closed, bounded from below and non-empty, there exists a smallest β with $h(\beta) = 0$. For this, the property $\phi(\beta') < \ell(\beta')$ necessarily holds for all $\beta' \in (0, \beta)$.

(ii) Since $0 = h(0) = h(\beta)$, it follows by the mean value theorem that there exists an $\alpha \in (0, \beta)$ such that

$$\phi'(\alpha) = \ell'(\alpha).$$

(iii) We thus obtain that for α.

$$\phi(\alpha) = f(x + \alpha d) < f(x) + c_1 \alpha \nabla f(x)^T d =: \ell(\alpha)$$

and because of $c_1 < c_2$

$$\phi'(\alpha) = \nabla f(x + \alpha d)^T d = c_1 \nabla f(x)^T d > c_2 \nabla f(x)^T d$$

holds.

Ex 4 **Programming** (12 Points)

Implement a function `gradientdescent(f, x, tol, maxit, method)` which takes as parameters a callable function `f`, and a numpy array `x`. The float and integer numbers `tol` and `maxit` will terminate the gradient descent algorithm after an error tolerance is reached or the maximum number of allowed iterations is exceeded. Implement the function such that it allows...

(i) to set `method="constant"`, which just evokes a gradient algorithm with constant step size $\alpha = 0.001$, which is not even a descent algorithm in general!

(ii) to set `method="armijo"`, which executes the *backtracking line-search*.

(iii) to set `method="wolfe"`, which executes a *wolfe line-search* by reducing the step size until the wolfe conditions are fulfilled.

(iv) Test your algorithm using the function `test(gradientdescent)` of the provided module `graddesc_test`.

Hint: You might like to incorporate a callback function as it is provided by the `class CallBack` in `graddesc_test`.