Math. Program., Ser. A
DOI 10.1007/s10107-015-0913-2

FULL LENGTH PAPER

Simple examples for the failure of Newton's method with line search for strictly convex minimization

Florian Jarre[1] · **Philippe L. Toint**[2]

Received: 2 November 2014 / Accepted: 13 May 2015

Abstract In this paper two simple examples of a twice continuously differentiable strictly convex function f are presented for which Newton's method with line search converges to a point where the gradient of f is not zero. The first example uses a line search based on the Wolfe conditions. For the second example, some strictly convex function f is defined as well as a sequence of descent directions for which exact line searches do not converge to the minimizer of f. Then f is perturbed such that these search directions coincide with the Newton directions for the perturbed function while leaving the exact line search invariant.

Keywords Newton's method · Line search · Wolfe conditions · Convex minimization

Mathematics Subject Classification Primary 90C25; Secondary 49M15

1 Introduction

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and let $\bar{x}$ be an accumulation point of the iterates generated by a descent method for f with a line search subject to the Wolfe conditions (shortly denoted by Wolfe line search). Then, under mild assumptions $\bar{x}$ is a stationary point, i.e. $\nabla f(\bar{x}) = 0$. When f is strictly convex and

✉ Philippe L. Toint
philippe.toint@unamur.be

Florian Jarre
jarre@hhu.de

1 Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

2 Namur Center for Complex Systems (naXys), University of Namur, 5000 Namur, Belgium

Published online: 24 May 2015

twice continuously differentiable, the Newton direction for finding a root of ∇f always is a descent direction whenever the Newton direction is well-defined. In this paper a simple example of a twice continuously differentiable strictly convex function f is presented which has a unique minimizer and for which Newton's method with a Wolfe line search converges to a point $\bar{x}$ with $\nabla f(\bar{x}) \neq 0$. The line search in this example is chosen as to avoid a certain set of "regular" points while meeting the Wolfe conditions. The convergence analysis is carried out for a well-chosen starting point, but is generalizable to other starting points as long as the line search can be manipulated to avoid the regular points. In a second example, a strictly convex function is constructed for a given starting point such that Newton's method with exact line search also converges to a non-stationary point. As far as the authors can see at this stage, this second example cannot be extended to general starting points.

In both examples, the Newton iterates $x^{(k)}$ converge to a non-stationary point $\bar{x}$ where the Hessian $\nabla^2 f(\bar{x})$ is singular, and the Newton directions $\frac{\Delta x^{(k)}}{\|\Delta x^{(k)}\|_2}$ with $\Delta x^{(k)} := -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$ have two limit directions at $\bar{x}$. Both of these limit directions are orthogonal to $\nabla f(\bar{x})$, and in particular, both are not descent directions for f at $\bar{x}$. This particular situation is crucial for constructing an example where Newton's method fails.

2 Known results on the convergence of Newton's method

We start by recalling in this section some results of Chapter 3.2 in [4] and some related results from [1–3]. Given a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, a point x, a direction Δx with $\nabla f(x)^T \Delta x < 0$, and constants $0 < c_1 < c_2 < 1$, a step length α is said to satisfy the Wolfe conditions if the following inequalities hold:

1. $f(x + \alpha \Delta x) \leq f(x) + c_1 \alpha \nabla f(x)^T \Delta x$
2. $\nabla f(x + \alpha \Delta x)^T \Delta x \geq c_2 \nabla f(x)^T \Delta x.$

Note that the set of points "$x + \alpha \Delta x$" satisfying the Wolfe conditions does not depend on the norm of Δx, i.e. setting $\Delta \tilde{x} := \mu \Delta x$ for some scalar $\mu > 0$, then $x + \alpha \Delta x$ satisfies the Wolfe conditions if, and only if, $x + \frac{\alpha}{\mu} \Delta \tilde{x}$ satisfies the Wolfe conditions. In the following example, the norm of the (full) Newton steps will grow unbounded and, at the same time, the norm of the Newton steps with Wolfe line search will go to zero.

A simple descent algorithm for minimizing f is given as follows:

Descent Algorithm:

1. Let some initial point $x^{(0)}$ be given. Let $\gamma \in (0, 1]$ and $0 < c_1 < c_2 < 1$ be given. Set $k := 0$.
2. If $\nabla f(x^{(k)}) = 0$, stop. Else choose $\Delta x^{(k)} \neq 0$ with $\nabla f(x^{(k)})^T \Delta x^{(k)} \leq -\gamma \|\nabla f(x^{(k)})\|_2 \, \|\Delta x^{(k)}\|_2$.
3. Set $x^{(k+1)} := x^{(k)} + \alpha_k \Delta x^{(k)}$ where α_k satisfies the Wolfe conditions.
4. Set $k := k + 1$ and go to Step 2.

If f is twice continuously differentiable and bounded below (i.e. $\exists M < \infty : \ f(x) \geq -M \ \forall x \in \mathbb{R}^n$), then the Wolfe conditions can be satisfied at every iteration k, and if the algorithm does not terminate after a finite number of steps it generates a sequence of iterates $\{x^{(k)}\}_k$, and each accumulation point x^* of this sequence is a critical point in the sense that $\nabla f(x^*) = 0$.

To discuss known results for the convex case, the following definitions will be used: For a point $x \in \mathbb{R}^n$ and $\epsilon > 0$, let $U_\epsilon(x) := \{z \mid \|x - z\|_2 < \epsilon\}$ denote the open ϵ-neighborhood of x. Let $S \subset \mathbb{R}^n$ be convex. A function $f : S \to \mathbb{R}$ is strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y) \quad \text{for all } \lambda \in (0,1)$$
$$\text{and all } x, y \in S \text{ with } x \neq y.$$

The function f is locally strongly convex at some point $x \in S$, if there exists $\epsilon(x) > 0$ such that the function $f_{\epsilon,x} : U_\epsilon(x) \cap S \to \mathbb{R}$ with $f_{\epsilon,x}(y) := f(y) - \epsilon(x)\|y\|_2^2$ is convex. It is locally strongly convex on S, if it is locally strongly convex at every $x \in S$. It is (globally) strongly convex on S, if there exists $\epsilon > 0$ independent of x such that $f_\epsilon := f(x) - \epsilon\|x\|_2^2$ is convex on S. Hence, $x \mapsto x^6$ is strictly convex but not locally strongly convex, and $x \mapsto e^x$ is locally strongly convex but not globally strongly convex.

When f is locally strongly convex and twice differentiable, the Newton step

$$\Delta x := -[\nabla^2 f(x)]^{-1} \nabla f(x)$$

for minimizing f satisfies $\nabla f(x)^T \Delta x < 0$. Moreover, if, in addition, f has a minimizer x^*, then Newton's method for minimizing f with a Wolfe line search globally converges to x^*. The same is true for Newton's method with exact line search. (This follows from the results in [4] when observing that the assumptions on f imply that the level set $\{x \mid f(x) \leq f(x^{(0)})\}$ is bounded, and hence, by Heine–Borel, the Hessian of f is uniformly positive definite on this set).

The convergence of the sequence of iterates of descent- and Newton-methods also has been considered in several recent papers. Absil et al. [1] show convergence of the sequence of the iterates of descent algorithms to a stationary point (i.e. only one accumulation point x^*) when f is analytic, an angle condition (Step 2. of the Descent Algorithm) is satisfied, and the Wolfe conditions hold (Theorem 4.1 (iii)). Byrd et al. [2] prove that Newton's method with line search cannot converge to a non-stationary point z when f is twice continuously differentiable, its Hessian is positive definite at all iterates, the Wolfe conditions hold, and $\nabla f(z) \notin \text{range}\nabla^2 f(z)$ where $\nabla^2 f(z)$ has rank $n-1$ (Theorem 3.2). They close with the statement "We have not been able to find an example of such false convergence when $\nabla f(z) \in \text{range}\nabla^2 f(z)$." This question has been answered by Mascarenhas [3] who uses highly oscillatory functions to construct an example of such false convergence. Mascarenhas also shows that any counter example must satisfy the condition that the step lengths α_k (subject to the Wolfe conditions) must tend to zero as $k \to \infty$. In this paper we are interested in the convex case and complement the result of Mascarenhas by presenting two further examples of false convergence where f is strictly convex.

In the next section we consider the case where $f : \mathbb{R}^n \to \mathbb{R}^n$ is twice continuously differentiable, has a unique minimizer x^*, and is strictly convex but not locally strongly convex. Moreover, f is locally strongly convex at almost all points x. More precisely, the points where f is not locally strongly convex form a set of measure zero that has

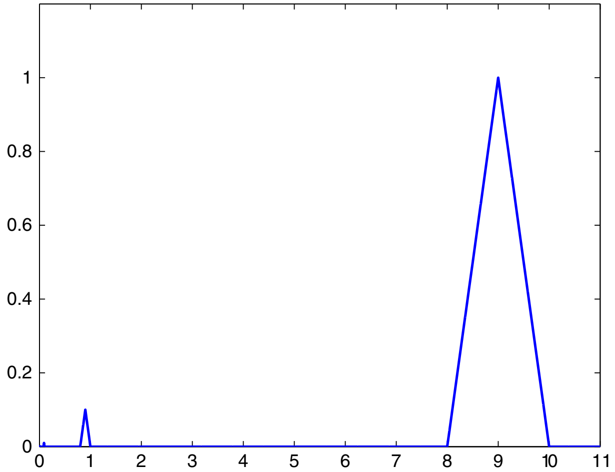

Fig. 1 The shape of the function $h_1^+(t)$

empty intersection with the iterates generated by the Newton algorithm. (In particular, all Newton directions are well defined and all Newton directions are descent directions for f at the current iterates—however, there does not exist a positive γ such that the condition in Step 2. of the Descent Algorithm is satisfied for all iterations).

3 A first example

Define a "hat-shaped-function" $\hat{h}_1 : \mathbb{R} \to \mathbb{R}$ via

$$\hat{h}_1(t) := \begin{cases} 0 & \text{for } t < 8 \\ t - 8 & \text{for } 8 \le t < 9 \\ 10 - t & \text{for } 9 \le t < 10 \\ 0 & \text{for } t \ge 10. \end{cases}$$

Then define a continuous function $h_1^+ : \mathbb{R} \to \mathbb{R}$ via

$$h_1^+(t) := \sum_{k=0}^{\infty} 10^{-k} \hat{h}_1 \left(10^k t \right).$$

The function h_1^+, illustrated in Fig. 1, is continuous and has infinitely many "hats" in the interval $(0, 10]$, where the height and width of the k-th "hat" tends to zero as $k \to \infty$.

More precisely, on each interval of the form $[10^{-k},\ 8 \cdot 10^{-k}]$ for $k = 0, 1, 2, \ldots$ and on $[10, \infty)$, the function h_1^+ is identically zero; for all other input arguments $t \in (0, 10)$ it is strictly positive. Integrating h_1^+ twice yields a convex function f_1^+ for which the second derivative is zero on each interval of the form $[10^{-k},\ 8 \cdot 10^{-k}]$. For the exact definition of f_1^+ let $g_1^+ : \mathbb{R} \to \mathbb{R}$ be defined via

$$g_1^+(t) := \int_0^t h_1^+(s)\, ds.$$

On each interval of the form $[10^{-k},\ 8 \cdot 10^{-k}]$ for $k = 0, 1, 2, \ldots$ the function g_1^+ is constant, and (since $\int_0^\infty 10^{-i} \hat{h}_1(10^i t)\, dt = 100^{-i}$)

$$g_1^+(10^{-k}) = \sum_{i=k+1}^{\infty} 100^{-i} = \frac{100^{-(k+1)}}{0.99} = \frac{1}{99} 10^{-2k}. \tag{1}$$

Let $f_1^+ : \mathbb{R} \to \mathbb{R}$ be defined via

$$f_1^+(t) := \int_0^t g_1^+(s)\, ds.$$

By construction, $f_1^+(t) = O(t^3)$ for $0 \le t \le 10$ (and $f_1^+(t) \equiv 0$ for $t \le 0$). Finally, let $f_1 : \mathbb{R} \to \mathbb{R}$, whose shape is shown in Fig. 2, be defined as $f_1(t) = f_1^+(t) + f_1^+(-t)$.

The function f_1 is convex, twice continuously differentiable, and satisfies a cubic growth condition near its unique minimizer $\bar{t} = 0$. Its derivatives are given by

$$f_1'(t) = g_1^+(t) - g_1^+(-t), \quad f_1''(t) = h_1^+(t) + h_1^+(-t).$$

A point t with $f_1''(t) > 0$ will be called a "regular" point. The following example is constructed as to avoid such regular points.

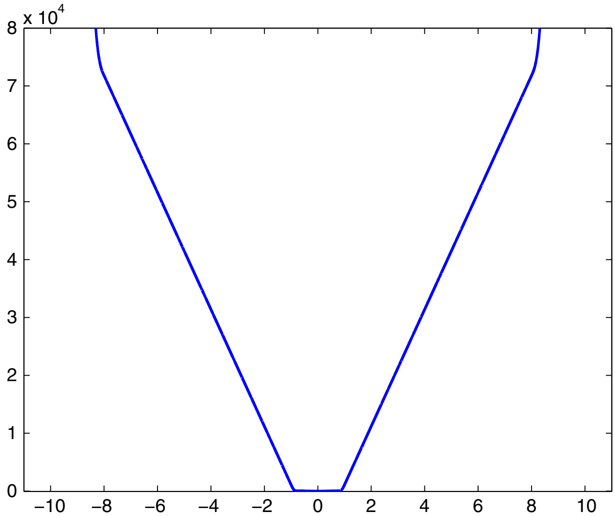

Fig. 2 The shape of the function $f_1(t)$

Let $\rho := \frac{1}{30 \cdot 90^2}$, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$f(x) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) := f_1(x_1) + \rho x_1^6 + x_2 + \frac{1}{2} x_2^2$$

The function f is strictly convex, twice continuously differentiable, and has a unique minimizer at $x^{min} := \begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

It is locally strongly convex at all points except from points x with $x_1 = 0$. Its first derivative is given by

$$\nabla f(x) = \left[f_1'(x_1) + 6\rho x_1^5,\ 1 + x_2\right]^T, \tag{2}$$

and its Hessian

$$\nabla^2 f(x) = \begin{bmatrix} f_1''(x_1) + 30\rho x_1^4 & 0 \\ 0 & 1 \end{bmatrix}$$

is positive definite at all points x except at points x with $x_1 = 0$.

The Newton direction for minimizing f at x with $x_1 \neq 0$ is given by

$$\Delta x := \begin{bmatrix} -\left(f_1'(x_1) + 6\rho x_1^5\right) / \left(f_1''(x_1) + 30\rho x_1^4\right) \\ -x_2 - 1 \end{bmatrix}.$$

Assume for the moment that an iterate $x^{(k)}$ of Newton's method is of the form $x^{(k)} := \begin{bmatrix} 10^{-k} \\ t \end{bmatrix}$ with $k \in \mathbb{N}_0$ and $t \in (0, \frac{1}{9}]$. In this case, by (1) and since $x^{(k)}$ is not "regular", the Newton direction simplifies to

$$\Delta x := \begin{bmatrix} -\left(\frac{1}{99} 10^{-2k} + 6\rho 10^{-5k}\right) / \left(30\rho 10^{-4k}\right) \\ -t - 1 \end{bmatrix}.$$

The numerator of the first component of Δx lies in the interval $[-\frac{1}{90} 10^{-2k}, -\frac{1}{100} 10^{-2k}]$ and the second component of Δx lies in the interval $[-\frac{10}{9}, -1]$. Hence the norm of Δx tends to infinity when $k \to \infty$ but, as detailed before, this does not influence the set of points that are acceptable for a line search along Δx based on the Wolfe conditions.

We now show that a point $x^{(k+1)}$ of the form $x^{(k+1)} = \begin{bmatrix} -10^{-k-1} \\ t \end{bmatrix}$ with $k \in \mathbb{N}_0$ and $t \in (0, \frac{1}{9}]$ satisfies the Wolfe conditions. Observe first that the above estimates on the components of Δx and (2) imply that

$$\begin{aligned}\nabla f\left(x^{(k)}\right)^T \Delta x &= \left(\frac{1}{99} 10^{-2k} + 6\rho 10^{-5k}\right) \Delta x_1 + (1 + t)\Delta x_2 \\ &\in \left[-\frac{\frac{1}{90^2} 10^{-4k}}{30\rho 10^{-4k}} - \left(\frac{10}{9}\right)^2,\ -\frac{\frac{1}{100^2} 10^{-4k}}{30\rho 10^{-4k}} - 1\right].\end{aligned}$$

Recalling that $\rho := \frac{1}{30 \cdot 90^2}$, this interval reduces to $[-\frac{181}{81}, -\frac{181}{100}]$. Likewise,

$$\nabla f\left(x^{(k+1)}\right)^T \Delta x \in \left[\frac{\frac{1}{100^2}10^{-4k-2}}{30\rho 10^{-4k}} - \left(\frac{10}{9}\right)^2, \frac{\frac{1}{90^2}10^{-4k-2}}{30\rho 10^{-4k}} - 1\right] \subset \left[-\frac{100}{81}, -\frac{99}{100}\right].$$

Hence, $\nabla f(x^{(k)})^T \Delta x \le -\frac{181}{100}$ and $\nabla f(x^{(k+1)})^T \Delta x \ge -\frac{100}{81}$. Since $\frac{100}{81}/\frac{181}{100} < 0.7$ the point $x^{(k+1)}$ satisfies the second Wolfe condition when c_2 is chosen $c_2 \in [0.7, 1)$. For points x on the line segment $[x^{(k)}, x^{(k+1)}]$ the above estimates imply that

$$\nabla f(x)^T \Delta x < \frac{99}{100}\frac{81}{181}\nabla f\left(x^{(k)}\right)^T \Delta x < 0.4 \nabla f\left(x^{(k)}\right)^T \Delta x$$

so that the first Wolfe condition is satisfied when c_1 is chosen $c_1 \in (0, 0.4]$.

When the initial point $x^{(0)}$ is chosen as $x^{(0)} := \begin{bmatrix} 1 \\ 1/9 \end{bmatrix}$, it is a simple exercise to verify that all iterates can be chosen of the form $x^{(k)} = \begin{bmatrix} \pm 10^{-k} \\ t \end{bmatrix}$ with $t \in (0, \frac{1}{9}]$. Observe now that the total length of the Newton path in the x_1-direction is less than 2. Observe also that the absolute value of the numerator of the first component of Δx is at least $\frac{1}{90 \cdot 10^{2k}}$ and the denominator is $\frac{1}{90^2 \cdot 10^{4k}}$, so that $|\Delta x_1| \ge 90$ for $k \ge 0$, while $|\Delta x_2| \le \frac{10}{9}$. As a result, we obtain that $|\Delta x_1| > 81\, |\Delta x_2|$. Thus, the x_2-component of the step it is always shorter by a factor at least 81 than its x_1 component, so that t converges to a number in the interval $(0, \frac{1}{9})$. The minimization algorithm therefore converges to a non-stationary point.

The above analysis depends on the choice of the Wolfe parameters $c_1 \le 0.4$ and $c_2 \ge 0.7$. Observe that the gradient of f at $x^{(0)}$ and the Newton direction at $x^{(0)}$ are continuous functions of $x^{(0)}$. Therefore, the Wolfe conditions at $x^{(1)} := \begin{bmatrix} -10^{-1} \\ t \end{bmatrix}$ with some $t \in (\frac{1}{9} - \delta, \frac{1}{9}]$ remain true for all choices of $x^{(0)}$ in an open set

$$x^{(0)} \in \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad | \quad x_1 \in (1, 1+\epsilon), \quad x_2 \in \left(\frac{1}{9} - \epsilon, \frac{1}{9}\right) \right\}$$

for some $\epsilon > 0$, when relaxing the bounds on c_1 and c_2 to, say, $c_1 \le 0.3$ and $c_2 \ge 0.8$. Here, $\delta > 0$ is some parameter with $\delta \to 0$ as $\epsilon \to 0$. When δ is sufficiently small, the preceeding analysis implies again that for $k \ge 1$ all iterates can be chosen of the form $x^{(k)} = \begin{bmatrix} \pm 10^{-k} \\ t \end{bmatrix}$ with $t \in (0, \frac{1}{9}]$. Thus, for all starting points in an open set the line search can be manipulated (subject to the Wolfe conditions) so that Newton's method converges to some non-stationary point. We believe that this property can also be maintained for many other starting points. The analysis does assume, however, that the line search can be manipulated in a way that "regular" points are avoided, i.e. only points are visited for which the second derivative of f_1 is zero. If an exact line search is used, this assumption is difficult to control. Nevertheless, as demonstrated in the next section, even an exact line search is not a guarantee that Newton's method converges to a minimizer.

4 Using exact line searches

To start this second example, consider the strictly convex and twice continuously differentiable function

$$\tilde{f} : \mathbb{R}^2 \to \mathbb{R}, \quad \tilde{f}(x) := |x_1|^3 + x_2 + \frac{1}{2}x_2^2. \tag{3}$$

having the same minimizer $x^{min} := \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ as f in Sect. 3

The Newton step for minimizing $\tilde{f}$ starting at some point x with $x_1 \neq 0$ is given by $\Delta x = \begin{bmatrix} -x_1/2 \\ -1 - x_2 \end{bmatrix}$, i.e. it is "too short" by a factor $1/2$ in the x_1-direction.

Consider, for the moment, the starting point $x^{(0)} := \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$ such that $\tilde{f}(x^{(0)}) = 0.106$ and a sequence of exact line search steps for minimizing $\tilde{f}$ along the search directions

$$\Delta x^{(k)} := \begin{bmatrix} (-1)^{k+1} \\ -\frac{1}{10^{2k+2}} \end{bmatrix},$$

each line search starting at a point $x^{(k)}$ and leading to a point $x^{(k+1)} := x^{(k)} + \alpha_k \Delta x^{(k)}$. Note that strict convexity of $\tilde{f}$ and the existence of a global minimizer imply that $x^{(k+1)}$ is well defined and unique. The exact coordinates $x_1^{(k+1)}$ and $x_2^{(k+1)}$ will not be analyzed in this paper, instead we derive lower and upper bounds for these coordinates. Because the sequence $\{\tilde{f}(x^{(k)})\}$ is monotonically decreasing by construction, we have that, for all k,

$$0.106 \geq \left|x_1^{(k)}\right|^3 + x_2^{(k)} + \frac{1}{2}\left(x_2^{(k)}\right)^2 \geq \left|x_1^{(k)}\right|^3 + \min_{x_2}\left[x_2 + \frac{1}{2}x_2^2\right] = \left|x_1^{(k)}\right|^3 - \frac{1}{2}$$

and thus $|x_1^{(k)}|^3 \leq 0.606$, implying $|x_1^{(k)}| \leq 0.85$. Since $|\Delta x_1^{(k)}| = 1$ for all k, this in turn ensures that $|\alpha_k| \leq 1.7$ for all k, and hence that

$$x_2^{(k+1)} \geq 0.1 - 1.7 \sum_{i=0}^{k} 10^{-2i-2} > 0.08 > 0$$

for $k > 0$. The exact line search implies further that

$$0 = \nabla \tilde{f}\left(x^{k+1}\right)^T \Delta x^{(k)} = 3(-1)^{k+1} \operatorname{sign}\left(x_1^{(k+1)}\right) \left(x_1^{(k+1)}\right)^2 \\ -10^{-2k-2}\left(1 + x_2^{(k+1)}\right),$$

and thus

$$\operatorname{sign}\left(x_1^{(k+1)}\right) = (-1)^{k+1} \quad \text{and} \quad x_1^{(k+1)} = (-1)^{k+1}\sqrt{\frac{1 + x_2^{(k+1)}}{3 \cdot 10^{2k+2}}}. \tag{4}$$

Because $0 < x_2^{(k+1)} \leq 0.1$, this implies that, for $k > 0$,

$$\left|x_1^{(k+1)}\right| \in \left[\frac{1}{10^{k+1}}\sqrt{\frac{1}{3}}, \frac{1}{10^{k+1}}\sqrt{\frac{1.1}{3}}\right],$$

resembling the situation of the example in the previous section. The limit of the sequence $x^{(k)}$ is a point $\bar{x} = \begin{bmatrix} 0 \\ \bar{x}_2 \end{bmatrix}$ with $\bar{x}_2 > 0$.

In the following the function $\tilde{f}$ shall be modified such that the above search directions coincide (up to positive multiples) with the Newton directions. This is achieved by "increasing" the x_1-component of the Newton step Δx for minimizing $\tilde{f}$, and this, in turn, is achieved by locally reducing $\frac{\partial^2}{\partial x_1^2}\tilde{f}(x^{(k)})$ while leaving $\nabla \tilde{f}(x^{(k)})$ invariant and while maintaining strict convexity of $\tilde{f}$. Convexity and unchanged first derivative at all points $x^{(k)}$ imply that the exact line searches are not affected by this modification.

We now define the local perturbations of $\tilde{f}$ using a B-spline. More precisely, let $s : \mathbb{R} \to \mathbb{R}$ be given by

$$s(t) := \begin{cases} 0 & \text{for } t < -2 \\ (t+2)^3 & \text{for } -2 \leq t < -1 \\ 1 + 3(t+1) + 3(t+1)^2 - 3(t+1)^3 & \text{for } -1 \leq t < 0 \\ 1 - 3(t-1) + 3(t-1)^2 + 3(t-1)^3 & \text{for } 0 \leq t < 1 \\ -(t-2)^3 & \text{for } 1 \leq t < 2 \\ 0 & \text{for } t \geq 2. \end{cases}$$

Its derivatives are given by

$$s'(t) = \begin{cases} 0 & \text{for } t < -2 \\ 3(t+2)^2 & \text{for } -2 \leq t < -1 \\ 3 + 6(t+1) - 9(t+1)^2 & \text{for } -1 \leq t < 0 \\ -3 + 6(t-1) + 9(t-1)^2 & \text{for } 0 \leq t < 1 \\ -3(t-2)^2 & \text{for } 1 \leq t < 2 \\ 0 & \text{for } t \geq 2, \end{cases}$$

$$s''(t) = \begin{cases} 0 & \text{for } t < -2 \\ 6(t+2) & \text{for } -2 \le t < -1 \\ 6 - 18(t+1) & \text{for } -1 \le t < 0 \\ 6 + 18(t-1) & \text{for } 0 \le t < 1 \\ -6(t-2) & \text{for } 1 \le t < 2 \\ 0 & \text{for } t \ge 2. \end{cases}$$

For $\tau \in \mathbb{R}\backslash\{0\}$ let $s_\tau : \mathbb{R} \to \mathbb{R}$, a scaled and shifted version of s, be defined via

$$s_\tau(t) := \frac{\tau^2}{192} s\left(\frac{4(t-\tau)}{\tau}\right).$$

By construction, the support of s_τ lies in the interval $(\frac{1}{2}\tau, \frac{3}{2}\tau)$ when $\tau > 0$ and in $(\frac{3}{2}\tau, \frac{1}{2}\tau)$ when $\tau < 0$, and the minimizer of s''_τ is at the point $t = \tau$ with

$$s'_\tau(\tau) = \frac{4}{\tau}\frac{\tau^2}{192}s'(0) = 0 \quad \text{and} \quad s''_\tau(\tau) = \frac{16}{\tau^2}\frac{\tau^2}{192}s''(0) = -1. \tag{5}$$

Now, consider a perturbation of $\tilde{f}_{\rho,\tau}$ of $\tilde{f}(x)$ defined by

$$\tilde{f}_{\rho,\tau}(x) := \tilde{f}(x) + \rho s_\tau(x_1)$$

for some parameters $\rho > 0$ and $\tau \in \mathbb{R}\backslash\{0\}$. Using (3), we obtain that

$$\nabla \tilde{f}_{\rho,\tau}(x) = \begin{bmatrix} 3\,\text{sign}(x_1)\,x_1^2 + \rho s'_\tau(x_1) \\ 1 + x_2 \end{bmatrix} \quad \text{and}$$
$$\nabla^2 \tilde{f}_{\rho,\tau}(x) = \begin{bmatrix} 6\,|x_1| + \rho s''_\tau(x_1) & 0 \\ 0 & 1 \end{bmatrix}. \tag{6}$$

To prove strict convexity of $f_{\rho,\tau}(x)$, it obviously suffices to show that $6\,|x_1| + \rho s''_\tau(x_1) > 0$ for all $x_1 \neq 0$. Sketching the piecewise linear graph for this quantity (see Fig. 3), it is easy to see that it is strictly positive for all $x_1 \neq 0$ if $6\,|\tau| + \rho s''_\tau(\tau) > 0$, i.e., in view of (5), if $\rho < 6\,|\tau|$.

In Fig. 3 the values of s''_τ are nonzero in the interval $x_1 \in [0.5, 1.5]$ and the lowest value of $6|x_1| + \rho s''_\tau(x_1)$ is at $x_1 = 1$.

Returning to the example at the beginning of this section, let $\tau_k = x_1^{(k)}$ for $k = 0, 1, 2, \ldots$ where $x_1^{(k)}$ is given in (4). Then, the supports of s_{τ_k} and of s_{τ_l} are disjoint for $k \neq l$ (compare with Fig. 3). Moreover, let

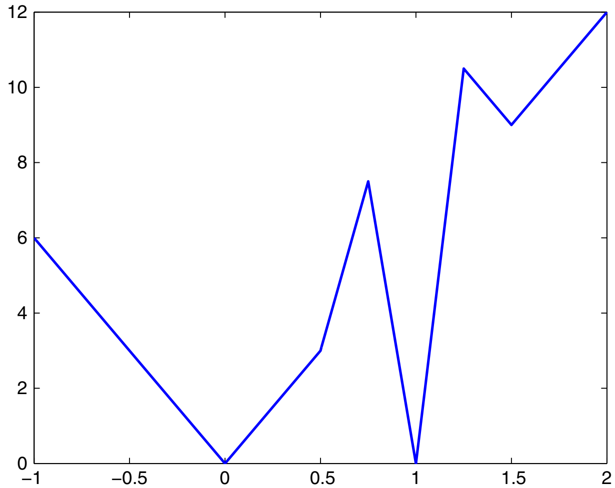

Fig. 3 The shape of "$6\,|x_1| + \rho s''_\tau(x_1)$" for $\tau = 1$ and $\rho = 6\tau$

$$\rho_k := 6\left|x_1^{(k)}\right| - 3 \cdot 10^{-2k-2}\frac{\left(x_1^{(k)}\right)^2}{1 + x_2^{(k)}} < 6\,|\tau_k|,$$

thereby ensuring the strict convexity of the function $\tilde{f}_{\rho_k, x_1^{(k)}}$. Moreover, we obtain from (5) and (6) that the Newton step from $x^{(k)}$ on $\tilde{f}_{\rho_k, x_1^{(k)}}$ is given by

$$-\left[\begin{array}{c}\left(3\,\mathrm{sign}\left(x_1^{(k)}\right)\,\left(x_1^{(k)}\right)^2\right) / \left(6\left|x_1^{(k)}\right| - \rho_k\right) \\ 1 + x_2^{(k)}\end{array}\right] = \frac{1 + x_2^{(k)}}{10^{-2k-2}}\begin{bmatrix}(-1)^{k+1} \\ -10^{-2k-2}\end{bmatrix},$$

which is a positive multiple of $\Delta x^{(k)}$. Moreover, the function f with $f(x) := \tilde{f}(x) + \sum_{k=0}^{\infty}\rho_k s_{\tau_k}(x_1)$ is well defined since at most one term in the sum is nonzero. That it is also twice continuously differentiable can be deduced from the same argument for $x_1 \neq 0$ and the fact that the boundedness of s_{τ_k}, s'_{τ_k} and s''_{τ_k} and the limit $\rho_k < 6|\tau_k| = 6|x_1^{(k)}| \to 0$ together ensure the desired property at $x_1 = 0$. (For the case $x_1 = 0$, again, the fact is exploited that the supports of s_{τ_k} and s_{τ_j} are disjoint for $k \neq j$). Thus, we have constructed a function f such that Newton's method with exact line search generates the same iterates as in (4) converging to a point where the first derivative is nonzero.

A necessary property for the above example to work is that the second derivative of $\tilde{f}$ at $\bar{x}$ is singular. Adding a term $\|x - \bar{x}\|_2^4$ to $\tilde{f}(x)$ will effect that $\bar{x}$ is the only point at which the second derivative of $\tilde{f}$ is singular. The somewhat tedious details for perturbing $\tilde{f}(\,.\,) + \|\,.\, - \bar{x}\|_2^4$ such that the line search still generates the same iterates have not been considered; it is conceivable, however, that the example can be modified such that f is locally strongly convex at all points except from $\bar{x}$.

Acknowledgments The authors are thankful to two anonymous referees for pointing out related references and for valuable criticism that helped improving the presentation of the paper.

References

1. Absil, P., Mahony, R., Andrews, B.: Convergence of the iterates of descent methods for analytic cost functions. SIAM J. Optim. **16**(2), 531–547 (2005)
2. Byrd, R., Marazzi, M., Nocedal, J.: On the convergence of Newton iterations to non-stationary points. Math. Program. Ser. A **99**, 127–148 (2004)
3. Mascarenhas, W.: Newton's iterates can converge to non-stationary points. Math. Program. Ser. A **112**, 327–334 (2008)
4. Wright, S.J., Nocedal, J.: Numerical Optimization, vol. 2. Springer, New York (2006)