Mathematical Programming manuscript No.
(will be inserted by the editor)

Roman A. Polyak

Regularized Newton Method for unconstrained Convex Optimization

Dedicated to B. T. Polyak on the occasion of his 70th birthday

Received: date / Revised version: date

Abstract. We introduce the regularized Newton method (rnm) for unconstrained convex optimization. For any convex function, with a bounded optimal set, the rnm generates a sequence that converges to the optimal set from any starting point. Moreover the rnm requires neither strong convexity nor smoothness properties in the entire space. If the function is strongly convex and smooth enough in the neighborhood of the solution then the rnm sequence converges to the unique solution with asymptotic quadratic rate. We characterized the neighborhood of the solution where the quadratic rate occurs.

1. Introduction.

There are several ways to modify the Newton method for unconstrained minimization to achieve global convergence.

For twice continuous differentiable and strongly convex function, the Newton direction is a descent direction. The local "quality" of the Newton direction at each point can be estimated by the condition number of the Hessian at the point.

If the condition number is bounded from above uniformly in $x \in \mathbf{R}^n$ then by introducing a step-size, it is possible to guarantee the global convergence of the so-called damped Newton method. By adjusting the step-size of the damped Newton method, using for example the Armijo rule, the asymptotic quadratic rate of convergence can be achieved.

To guarantee the global convergence of the Newton method in case when the function is not strongly convex, the Levenberg-Marquardt regularization of the Hessian is used (see [3], [4]).

In [5], Yu. Nesterov and A. Nemirovski introduced the class of self-concordant functions. These are three time differentiable convex function with the second and third derivatives satisfying a particular condition at each point.

The choice of the step-size is based on the value of the so-called Newton's decrement. It guarantees global convergence allowing estimation of the complexity, i.e. finding the upper bound for the number of iterations required to achieve the desired accuracy. To compute the Newton's decrement at each iteration, one has to invert the Hessian.

Roman A. Polyak: Department of SEOR and Mathematical Sciences Department, George Mason University, 4400 University Dr, Fairfax VA 22030, rpolyak@gmu.edu

Mathematics Subject Classification (1991): 20E28, 20G40, 20C20

Recently, Yu. Nesterov and B. Polyak proposed in [6] an interesting cubic regularization of the Newton method. At each iteration, it requires solving an unconstrained minimization problem with the same quadratic term as in the Newton method, but regularized via a cubic term.

Usually the convergence results for the Newton method include assumptions on a starting point which has to be in the neighborhood of the solution. In contrast, S. Smale [11] established convergence based on the assumptions on data at the point. The assumptions, however, include the existence and the boundness of the derivatives of any order at the point.

All mentioned modifications of the Newton method, as well as the classical Newton method (see L. Kantorovich [2]), require the existence and the continuity of the Hessian and its inverse.

Our main motivation is to develop a Newton type method for finding a minimum of a convex function that requires neither strong convexity nor even smoothness properties on the entire space.

We introduce and analyze the regularized Newton method (rnm) which can be applied for finding a minimum of any convex function $f : \mathbf{R}^n \to \mathbf{R}$ from any starting point $x \in \mathbf{R}^n$. The regularization of the convex function at each point with the norm of the gradient (subgradient) as the regularization parameter is the main idea behind the rnm.

The following regularized at the point x function:

$$F(x, y) = f(y) + \frac{1}{2}||\nabla f(x)||||y - x||^2$$

is our main tool.

Such regularization allows developing a Newton-type method, which generates a sequence that converges to the $\min_{x \in \mathbf{R}^n} f(x)$ for any convex function from any starting point $x \in \mathbf{R}^n$.

Moreover, the rnm retains the main property of the Newton method – the asymptotic quadratic rate of convergence for functions that satisfy the standard for the Newton method assumptions on $f : \mathbf{R}^n \to \mathbf{R}$ in the neighborhood of the solution. The size of the neighborhood of the solution where the quadratic convergence occurs is characterized through the convexity constant of $f(x)$ and the Lipschitz constant of its Hessian $\nabla^2 f(x)$ as well as a parameter $0 < r < 1$, which defines the quadratic rate.

The paper is organized as follows. The regularized convex function and the regularized Newton direction (rnd) are introduced in Section 2. The "quality" of the rnd as well as the "quality" of classical Newton direction (cnd) are quantified. We use these characteristics to show that regularization improves the condition number of the Hessian $\nabla^2 f(x)$, $\forall x \in \mathbf{R}^n, x \neq x^*$. The local quadratic convergence of the rnm sequence is proven in Section 3. The damped regularized Newton methods are introduced in Section 4. The global convergence and the convergence rate of the damped regularized Newton methods are established in Section 5.

2. Regularization of a convex function at the point.

Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function. We assume that the optimal set

$$X^* = \operatorname{Arg\,min}\{f(x) | x \in \mathbf{R}^n\} \tag{1}$$

is not empty and bounded.

In this section, we introduce the regularized convex function at the point and define the regularized Newton direction (rnd) for any convex function. We show that the regularized Newton direction exists at each $x \notin X^*$ no matter the function is strongly convex and smooth or not. Moreover, the regularization improves the condition number of the Hessian $\nabla^2 f(x)$ uniformly in $x \in \mathbf{R}^n$, $x \neq x^*$.

We assume at the beginning that $f \in C^2$. The Euclidean norm $\| x \| = (x, x)^{\frac{1}{2}}$ is used throughout the paper. The regularized, at the point $x \in \mathbf{R}^n$, function $f(x)$ we define by the following formula

$$F(x, y) = f(y) + \frac{1}{2} \| \nabla f(x) \| \| y - x \|^2 . \tag{2}$$

For any $x \notin X^*$ we have $||\nabla f(x)|| > 0$, therefore for any convex function $f(x)$ the regularized function $F(x, y)$ is strongly convex in $y \in \mathbf{R}^n$. Therefore for any $x \in \mathbf{R}^n$ there exists a unique minimizer

$$y(x) = \arg\min\{F(x, y) | y \in \mathbf{R}^n\}.$$

The following properties of the function $F(x, y)$ are direct consequences of (2).
1°. $F(x, y)_{|y=x} = f(x)$,
2°. $\nabla_y F(x, y)_{|y=x} = \nabla f(x)$,
3°. $\nabla^2_{yy} F(x, y)_{|y=x} = \nabla^2 f(x) + ||\nabla f(x)|| I = H(x)$, where I is the identity matrix in $\mathbf{R}^n$.

For any $x \notin X^*$, the inverse $H^{-1}(x)$ exists whether the function $f(x)$ is strongly convex or not. Therefore the regularized Newton step

$$\hat{x} := x - (H(x))^{-1} \nabla f(x) \tag{3}$$

can be performed from any starting point $x \notin X^*$ for any smooth but not necessarily strongly convex function $f(x)$. Moreover, we will see later that for any convex function $f(x)$, even the smoothness is not necessary to perform the regularized Newton step.

We start by showing that the regularization (2) improves the condition number of the Hessian $\nabla^2 f(x), \forall x \in \mathbf{R}^n, x \neq x^*$. We assume at this point that for any given $x \in \mathbf{R}^n$ there exist $0 < m(x) < M(x) < \infty$ such that

$$m(x)||y||^2 \leq (\nabla^2 f(x) y, y) \leq M(x)||y||^2, \qquad \forall y \in \mathbf{R}^n. \tag{4}$$

Along with the regularized Newton step (3), we consider the classical Newton step

$$\hat{x} := x - (\nabla^2 f(x))^{-1} \nabla f(x). \tag{5}$$

Let's quantify the "quality" of the regularized Newton direction $r(x)$, defined by the system

$$H(x)r(x) = -\nabla f(x) \tag{6}$$

and compare it with the correspondent characteristic of the cnd

$$n(x) = -(\nabla^2 f(x))^{-1}\nabla f(x). \tag{7}$$

At any point $x \in \mathbf{R}^n$, the local "quality" of a given descent direction $d \in R^n$ is characterized by the following number

$$0 \le q(d) = -\frac{(\nabla f(x), d)}{||\nabla f(x)||||d||} \le 1.$$

It is well-known that the steepest descent direction $d(x) = -\nabla f(x) \parallel f(x) \parallel = \arg\max\{q(d) | \parallel d \parallel \le 1\}$ is the best local descent direction and $q(d(x)) = 1$.

For the rnd $r(x)$ we have

$$q(r(x)) = -\frac{(\nabla f(x), r(x))}{||\nabla f(x)||||r(x)||}.$$

From (6), we obtain

$$(H(x)r(x), r(x)) = -(\nabla f(x), r(x)).$$

Keeping in mind 3° and the left inequality in (4), we obtain

$$-(\nabla f(x), r(x)) \ge (m(x) + ||\nabla f(x)||)||r(x)||^2.$$

Therefore, for any $x \notin X^*$ we have

$$1 \ge q(r(x)) \ge (m(x) + ||\nabla f(x)||)||r(x)||||\nabla f(x)||^{-1} > 0. \tag{8}$$

In other words, for any $x \notin X^*$, the rnd $r(x)$ is a descent direction for $f(x)$ no matter the function $f(x)$ is strongly convex or not. Whereas the cnd $n(x)$ exists and it is a descent direction only for a strongly convex function. The local "quality" of the cnd $n(x)$ can be characterized by

$$q(n(x)) = -\frac{(\nabla f(x), n(x))}{||\nabla f(x)||||n(x)||}.$$

The following theorem establishes the lower bounds for $q(r(x))$ and $q(n(x))$ and shows that the regularization (2) improves the condition number of the Hessian $\nabla^2 f(x)$ for all $x \in \mathbf{R}^n, x \ne x^*$.

Theorem 1. *Let $f \in C^2$ be a convex function that satisfy (4), then the following bounds hold*

1.

$$1 \ge q(r(x)) \ge (m(x) + ||\nabla f(x)||)(M(x) + ||\nabla f||)^{-1} = (cond\, H(x))^{-1} > 0, \quad \forall x \notin X^*.$$

2.

$$1 \geq q(n(x)) \geq m(x)(M(x))^{-1} = (cond\, \nabla^2 f(x))^{-1} > 0, \;\; \forall x \in \mathbf{R}^n : m(x) > 0.$$

3.

$$cond\, \nabla^2 f(x) - cond\, H(x) = ||\nabla f(x)||(cond\, \nabla^2 f(x) - 1)(m(x) + ||\nabla f(x)||)^{-1} > 0,$$
$$\forall x \notin X^*, cond\, \nabla^2 f(x) \neq 1$$

[Proof]:

1. From (6), we obtain

$$||\nabla f(x)|| \leq ||H(x)||||r(x)||. \tag{9}$$

Using the right inequality (4) and 3°, we have

$$||H(x)|| \leq M(x) + ||\nabla f(x)||, \tag{10}$$

From (9) and (10) we obtain

$$||\nabla f(x)|| \leq (M(x) + ||\nabla f(x)||)||r(x)||.$$

Combining the last inequality with (8) we have

$$q(r(x)) \geq (m(x) + ||\nabla f(x)||)(M(x) + ||\nabla f(x)||)^{-1} = (\text{cond } H(x))^{-1}.$$

2. Now let's consider the Newton direction $n(x)$. From (7), we have

$$\nabla f(x) = -\nabla^2 f(x) n(x), \tag{11}$$

therefore,

$$-(\nabla f(x), n(x)) = (\nabla^2 f(x) n(x), n(x)).$$

Hence, using the left inequality of (4), we have

$$q(n(x)) = -\frac{(\nabla f(x), n(x))}{||\nabla f(x)||||n(x)||} \geq m(x)||n(x)||||\nabla f(x)||^{-1}. \tag{12}$$

From (11) and the right inequality in (4), we obtain

$$||\nabla f(x)|| \leq ||\nabla^2 f(x)||||n(x)|| \leq M(x)||n(x)||. \tag{13}$$

Combining (12) and (13) we have

$$q(n(x)) \geq \frac{m(x)}{M(x)} = (\text{cond } \nabla^2 f(x))^{-1}.$$

3. Using the formulas for the condition numbers of $\nabla^2 f(x)$ and $H(x)$ we obtain

$$\begin{aligned}\text{cond } \nabla^2 f(x) - \text{cond } H(x) &= M(x)m(x)^{-1} - (M(x) + ||\nabla f(x)||)(m(x) + ||\nabla f(x)||)^{-1} \\ &= ||\nabla f(x)||(\text{cond } \nabla^2 f(x) - 1)(m(x) + ||\nabla f(x)||)^{-1} \\ &> 0, \qquad \forall x \notin X^*, \text{cond } \nabla^2 f(x) \neq 1\end{aligned}$$

□

It follows from Theorem 1 that the regularization (2) improves the condition number of the Hessian for any strongly convex function $f(x)$ uniformly in $x \in \mathbf{R}^n, x \neq x^*$, cond $\nabla^2 f(x) \neq 1$. It also makes possible to perform rnm when the original function is not strongly convex. Moreover, we will see later that the rnd exists even if $f(x)$ is not smooth. At the same time, the rnm retains the quadratic rate of convergence in the neighborhood of the solution.

We discuss this issue in the next section.

3. Local Regularized Newton method.

In this section, the regularized Newton method (rnm) is considered for smooth and strongly convex functions. It is shown that rnm generates a sequence that converges to the solution with quadratic rate from any starting point $x \in S(x^*, \rho) = \{x : ||x - x^*|| \leq \rho\}$.

We estimate $\rho > 0$ through the convexity constant of the function $f(x)$ and the Lipschitz constant of its Hessian $\nabla^2 f(x)$, as well as the parameter $0 < r < 1$, which characterize the quadratic convergence rate.

One step of the rnm consists of finding the approximation $\hat{x}$ by the following formula

$$\hat{x} := x - (\nabla^2 f(x) + ||\nabla f(x)||I)^{-1} \nabla f(x). \tag{14}$$

Along with rnm we consider a step of the classical Newton method (cnm)

$$\hat{x} := x - (\nabla^2 f(x))^{-1} \nabla f(x). \tag{15}$$

We assume that in the neighborhood $S(x^*, \rho_0)$ the Hessian $\nabla^2 f(x)$ satisfies the following standard for the cnm conditions

$$||\nabla^2 f(x) - \nabla^2 f(y)|| \leq L||x - y|| \tag{16}$$

$$m(y, y) \leq (\nabla^2 f(x)y, y) \leq M(y, y) \tag{17}$$

and $0 < m < M < \infty$.

The following inequalities are direct consequence of the conditions (16) and (17) (see [9])

$$||\nabla f(x + \triangle x) - \nabla f(x)|| \leq M||\triangle x|| \tag{18}$$

$$||\nabla f(x + \triangle x) - \nabla f(x) - \nabla^2 f(x)\triangle x|| \leq \frac{1}{2} L||\triangle x||^2. \tag{19}$$

In fact, in view of $f(x) \in C^2$, we have

$$\nabla f(x+\triangle x) = \nabla f(x) + \int_0^1 \nabla^2 f(x+\tau\triangle x)\triangle x d\tau \tag{20}$$

Using (20), we obtain

$$\nabla f(x+\triangle x) = \nabla f(x) + \nabla^2 f(x)\triangle x + \int_0^1 \left(\nabla^2 f(x+\tau\triangle x) - \nabla^2 f(x)\right)\triangle x d\tau.$$

Keeping in mind (16) we have

$$||\nabla f(x+\triangle x) - \nabla f(x) - \nabla^2 f(x)\triangle x|| \le L||\triangle x||^2 \int_0^1 \tau d\tau = \frac{1}{2}L||\triangle x||^2.$$

The next theorem establishes the quadratic rate of convergence of the rnm.

Theorem 2. *If conditions (16) and (17) are satisfied, then for any given*

$$0 < r = (m+0.5L)m^{-2}||\nabla f(x^0)|| < 1, x^0 \in S(x^*, \rho_0)$$

and

$$0 < \rho = m(m+0.5L)^{-1}r \le \rho_0 \tag{21}$$

the sequence $\{x^s\}_{s=0}^{\infty}$ *generated by the rnm (14) belongs to* $S(x^*, \rho)$ *and the following bound holds*

$$||x^s - x^*|| \le \frac{2m}{L+2m} r^{2^s}, \quad s \ge 1 \tag{22}$$

[Proof]: We consider the rnm step (14). Let $\triangle x = -(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}\nabla f(x)$, then using (19) we obtain

$$\begin{aligned} &||\nabla f(\hat{x}) - \nabla f(x) - \nabla^2 f(x)\triangle x|| \le 0.5L||\triangle x||^2 \\ &\le 0.5L||\left(\nabla^2 f(x) + ||\nabla f(x)||I\right)^{-2}|| \, ||\nabla f(x)||^2. \end{aligned} \tag{23}$$

For the vector $u = \nabla f(x) + \nabla^2 f(x)\triangle x$, we have

$$\begin{aligned} ||u|| &= ||\nabla f(x) + \nabla^2 f(x)\triangle x|| \\ &= ||\nabla f(x) - \nabla^2 f(x)(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}\nabla f(x)|| \\ &= ||\left(I - \nabla^2 f(x)(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}\right)\nabla f(x)|| \\ &\le ||I - \nabla^2 f(x)(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}|| \, ||\nabla f(x)||. \end{aligned}$$

Then

$$\begin{aligned} &||I - \nabla^2 f(x)(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}|| \\ &\qquad = ||(\nabla^2 f(x) + ||\nabla f(x)||I - \nabla^2 f(x))|| \, ||(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}||. \end{aligned}$$

From (17) we obtain $||(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}|| \le (m + ||\nabla f(x)||)^{-1}$ and therefore

$$||u|| \le (m + ||\nabla f(x)||)^{-1}||\nabla f(x)||^2. \tag{24}$$

From (23) for the vector $v = \nabla f(\hat{x}) - u$, we have

$$||v|| = ||\nabla f(\hat{x}) - u|| \leq 0.5L||\triangle x||^2. \tag{25}$$

Keeping in mind (24), (25) and

$$\begin{aligned} ||\triangle x|| &\leq ||(\nabla^2 f(x) + ||\nabla f(x)||)^{-1}||\ ||\nabla f(x)|| \\ &\leq (m + ||\nabla f(x)||)^{-1}||\nabla f(x)|| \end{aligned}$$

we obtain

$$\begin{aligned} ||\nabla f(\hat{x})|| &= ||u + v|| \\ &\leq ||u|| + ||v|| \\ &\leq \left((m + ||\nabla f(x)||)^{-1} + 0.5L(m + ||\nabla f(x)||)^{-2}\right)||\nabla f(x)||^2 \\ &\leq (m^{-1} + 0.5Lm^{-2})||\nabla f(x)||^2. \end{aligned}$$

Therefore for the rnm sequence $\{x^s\}$ generated by (14) we have

$$||\nabla f(x^{s+1})|| \leq \frac{m + 0.5L}{m^2}||\nabla f(x^s)||^2$$

or

$$\frac{m + 0.5L}{m^2}||\nabla f(x^{s+1})|| \leq \left(\frac{m + 0.5L}{m^2}||\nabla f(x^s)||\right)^2. \tag{26}$$

Let $r = \frac{m+0.5L}{m^2}||\nabla f(x^0)|| < 1$, then from (26) we obtain

$$||\nabla f(x^s)|| \leq \frac{m^2}{m + 0.5L}r^{2^s}. \tag{27}$$

It follows from left inequality (17) that

$$||\nabla f(x^0)|| \geq m||x^0 - x^*||. \tag{28}$$

Also from $||\nabla f(x^0)|| = \frac{m^2}{m+0.5L}r$ and (28) we have

$$||x^0 - x^*|| \leq m^{-1}||\nabla f(x^0)|| = \frac{mr}{m + 0.5L} = \rho < \rho_0$$

i.e. $x^0 \in S(x^*, \rho)$

Again using the left inequality (17) and (27), we obtain

$$||x^s - x^*|| \leq \frac{m}{m + 0.5L}r^{2^s}.$$

i.e. $x^s \in S(x^*, \rho), s \geq 1$ and the rnm sequence $\{x^s\}_{s=0}^{\infty}$ converges to x^* with quadratic rate. □

4. Global Regularized Newton methods.

To guarantee the global convergence of the cnm, the Newton direction $n(x)$ is used with a step length $t > 0$, i.e.

$$\hat{x} := x - t(\nabla^2 f(x))^{-1}\nabla f(x) = x + tn(x). \tag{29}$$

The step length t is often chosen from the Armijo inequality

$$f(x + tn(x)) \leq f(x) + ct(\nabla f(x), n(x)) \tag{30}$$

with $0 < c < 0.5$ (see [9]).

To guarantee the global convergence of the cnm with step length (30) one has to assume that $m(x) \geq m > 0$ and $0 < cond\, \nabla^2 f(x) < \infty$ for all $x \in \mathbf{R}^n$. It can be guaranteed by assuming that (16)–(17) takes place for all $x \in \mathbf{R}^n$. In other words the function $f(x)$ has to be strongly convex and smooth enough on $\mathbf{R}^n$.

In this section we consider two regularized Newton methods with step length. The first

$$x := x - t(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}\nabla f(x) \tag{31}$$

is designed for convex functions $f \in C^2$, which are not strongly convex on $\mathbf{R}^n$. For such class of convex functions the cnm cannot be applied ($m(x) = 0$) or it is impossible to guarantee convergence from any starting point. For example, for $f(x) = (1 + x^2)^{\frac{1}{2}}$ the cnm diverges for any starting point $x \notin (-1, 1)$.

The second rnm is designed for convex functions, which are neither strongly convex nor even differentiable in the entire $\mathbf{R}^n$. For both rnm we establish convergence from any starting point and in case when the conditions (16)–(17) are satisfied then both rnm converge with asymptotic quadratic rate.

We start with $f \in C^2$. From boundness of X^* follows that for any given $x^0 \in \mathbf{R}^n$ the closed convex set $\Omega = \{x : f(x) \leq f(x^0)\}$ is bounded. Therefore, there are $0 < L_0 = \max\{||\nabla^2 f(x)|| \mid x \in \Omega\}$ and $0 < M_0 = \max\{||\nabla f(x)|| \mid x \in \Omega\}$. Also for any pair $(x; y) \in \Omega \times \Omega$ we have

$$||\nabla f(x) - \nabla f(y)|| \leq L_0||x - y|| \tag{32}$$

Let Y be a closed bounded set, $x \notin Y$ and $d(x, Y) = ||x - y(x)|| = \min\{||x - y|| \mid y \in Y\}$ is the distance from x and Y.

The following theorem holds.

Theorem 3. *If $f \in C^2$ then the rnm (31) with step length $t = t(x) = (m(x) + ||\nabla f(x)||)L_0^{-1}$ generates a sequence $\{x^s\}_{s=0}^{\infty}$ that*

1. $\lim\limits_{s\to\infty} ||\nabla f(x^s)|| = 0$; *2.* $\lim\limits_{s\to\infty} f(x^s) = f(x^*)$; *3.* $\lim\limits_{s\to\infty} d(x^s, X^*) = 0$.

[Proof]: Let's consider the rnm (31)

$$\hat{x} := x + tr(x),$$

where $r(x) = -(\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}\nabla f(x)$.

We remind that for a pair of vectors $(x; y) \in \mathbf{R}^n \times \mathbf{R}^n$ we have

$$f(x+y) = f(x) + \int_0^1 (\nabla f(x+\tau y), y)\, d\tau \tag{33}$$

Using (33) we obtain

$$\begin{aligned} f(\hat{x}) &= f(x) + t(\nabla f(x), r(x)) + t\int_0^1 (\nabla f(x+\tau t r(x)) - \nabla f(x), r(x))\, d\tau \\ &\le f(x) + t(\nabla f(x), r(x)) + t\int_0^1 ||\nabla f(x+\tau t r(x)) - \nabla f(x)|| \cdot ||r(x)|| d\tau \end{aligned}$$

Using (6) and (32) we obtain

$$\begin{aligned} f(\hat{x}) &\le f(x) - t\left(\left(\nabla^2 f(x) + ||\nabla f(x)||I\right) r(x), r(x)\right) + 0.5t^2 L_0 ||r(x)||^2 \\ &\le f(x) - \left[t(m(x) + ||\nabla f(x)||) - 0.5t^2 L_0\right] ||r(x)||^2 \end{aligned}$$

Therefore for $t = t(x) = (m(x) + ||\nabla f(x)||)L_0^{-1}$, we have

$$f(\hat{x}) \le f(x) - 0.5L_0^{-1}(m(x) + ||\nabla f(x)||)^2 ||r(x)||^2 \tag{34}$$

From $r(x) = (\nabla^2 f(x) + ||\nabla f(x)||I)^{-1}\nabla f(x)$, we obtain

$$\begin{aligned} ||\nabla f(x)|| &\le (||\nabla^2 f(x)|| + ||\nabla f(x)||)||r(x)|| \\ &\text{or} \\ ||r(x)|| &\ge \frac{||\nabla f(x)||}{||\nabla^2 f(x)|| + ||\nabla f(x)||} \ge \frac{||\nabla f(x)||}{L_0 + ||\nabla f(x)||} \end{aligned}$$

Combining the last inequality with (34) and keeping in mind $m(x) \ge 0$ and $||\nabla f(x)|| \le M_0$ we obtain

$$\begin{aligned} f(\hat{x}) &\le f(x) - 0.5L_0^{-1}\left(\frac{m(x) + ||\nabla f(x)||}{L_0 + ||\nabla f(x)||}\right)^2 ||\nabla f(x)||^2 \\ &\le f(x) - 0.5L_0^{-1}(L_0 + M_0)^{-2}||\nabla f(x)||^4 \end{aligned}$$

Therefore the rnm (31) with step length

$$t = t(x) = (m(x) + ||\nabla f(x)||)L_0^{-1} \tag{35}$$

generates a sequence $\{x^s\}_{s=0}^{\infty}$ that

$$f(x^{s+1}) \le f(x^s) - 0.5L_0^{-1}(L_0 + M_0)^{-2}||\nabla f(x^s)||^4 \tag{36}$$

By summing up (36) we obtain

$$0.5L_0^{-1}(L_0+M_0)^{-2}\sum_{s=0}^{\infty}||\nabla f(x^s)||^4 \le f(x^0)-f(x^*)$$

Therefore

$$\lim_{s\to\infty}||\nabla f(x^s)||=0 \tag{37}$$

The sequence $\{x^s\}_{s=0}^{\infty}\subset\Omega$ is bounded. Let $\{x^{s_i}\}_{i=1}^{\infty}\subset\{x^s\}_{s=0}^{\infty}$ be a converging subsequence that $\lim_{s\to\infty}x^{s_i}=\bar{x}$. From (37) we have $\nabla f(\bar{x})=0$ and $\bar{x}=x^*\in X^*$. Also $\{f(x^s)\}_{s=0}^{\infty}$ is monotone decreasing. Therefore

$$\lim_{s\to\infty}f(x^s)=f(x^*),\quad\text{and}\quad\lim_{s\to\infty}d(x^s,X^*)=0\,.$$

□

The choice of the step length $t=t(x)$ requires knowledge of $L_0>0$ or its upper bound. Sometimes such upper bound can be found a priori. If it is not the case one can start with some $L_0>0$ and adjust its value if necessary during the solution process. It can be done similar to the way the correspondent parameter is adjusted in the second global regularized Newton method (grnm) which we describe later.

Now we can formulate the first grnm for convex $f\in C^2$, assuming that $L_0>0$ is available.

First Global Regularized Newton Method

Initialization: Given accuracy $\epsilon>0$, large enough $L>0$ and a starting point $x^0\in\mathbf{R}^n$
Set $\hat{x}:=x^0$.
step 1: $x:=\hat{x}$, if $||\nabla f(x)||\le\epsilon$, then stop and output $x^*:=x$.
step 2: Find $r(x)$ from (6), set $t:=1$, and find $\hat{x}$ from (31).
step 3: If $f(\hat{x})\ge f(x)$, then go to step 5.
step 4: If $||\nabla f(\hat{x})||\le||\nabla f(x)||^{1.5}$, then go to step 1.
step 5: Set $t:=t(x)$. Find $\hat{x}$ from (31) and go to step 1.

If the function $f\in C^2$ is convex but not strongly convex in $\mathbf{R}^n$, then the global convergence of the first grnm follows directly from Theorem 3.

If in the neighborhood $S(x^*,\rho)$ conditions (16)–(17) are satisfied, then the global convergence of the first grnm with asymptotic quadratic rate is a direct consequence of Theorem 2 and 3.

Let us estimate the number of steps of the regularized Newton method required to get an approximation with a given accuracy $\epsilon\ll\rho_0$ from any starting point $x^0\in\mathbf{R}^n$. It follows from (36) that

$$0.5L_0^{-1}(L_0+M_0)^{-2}\sum_{s=0}^{N_0}\|\nabla f(x^s)\|^4 \le f(x^0)-f(x^*)\,.$$

Let

$$\left\|\nabla f^*_{N_0}\right\| = \min_{0\le s\le N_0} \|\nabla f(x^s)\| \,,$$

then,

$$N_0 \left\|\nabla f^*_{N_0}\right\|^4 \le 2L_0(L_0+M_0)^2(f(x^0)-f(x^*))$$

and

$$\left\|\nabla f^*_{N_0}\right\| \le \sqrt[4]{\left(\frac{2L_0(L_0+M_0)^2(f(x^0)-f(x^*))}{N_0}\right)}\,.$$

It follows from Theorem 2 that the quadratic convergence (22) is taking place from any $x \in S(x^*,\rho_0)$, i.e.,

$$x : \|\nabla f(x)\| \le \frac{m^2 r}{m+0.5L}\,.$$

Therefore, if

$$\sqrt[4]{\frac{2L_0(L_0+M_0)^2(f(x^0)-f(x^*))}{N_0}} \le \frac{m^2 r}{m+0.5L}$$

then $x^{N_0} \in S(x^*,\rho_0)$ and for a given $\epsilon \ll \rho_0$ to get an approximation $x : \|x-x^*\| \le \epsilon$ requires $O(\ln\ln\epsilon^{-1})$ Newton steps, if x^{N_0} is taken as the starting point. Therefore, the total number of steps is

$$N = N_0 + O(\ln\ln\epsilon^{-1})\,, \tag{38}$$

where $N_0 = 2(m+0.5L)^4 m^{-8} r^{-2}(L_0+M_0)^2 L_0(f(x^0)-f(x^*))$.

We would like to point out that the number of Newton steps N is independent of the size of the problem.

Example 1. We consider $f(x) = (1+x^2)^{1/2}$, then $f'(x) = x(1+x^2)^{-(1/2)}$, $f''(x) = (1+x^2)^{-(3/2)}$ and $\min_{x\in R} f(x) = f(0) = 1$.

The Newton iteration is given by the following formula:

$$\hat{x} = x - (f''(x))^{-1} f'(x) = -x^3\,.$$

Therefore, the Newton method converges only from $x \in (-1,1)$.

Now, we consider the first grnm (31) with $t = t(x) = (m(x)+\|\nabla f(x)\|)L_0^{-1}$. We have $L_0 = \sup_{x\in R} f''(x) = 1$, $m(x) = f''(x)$, $\|\nabla f(x)\| = |f'(x)| = |x|(1+x^2)^{-(1/2)}$. Therefore, the first grnm iteration is given by the following formula:

$$\hat{x} = x - \frac{x}{\sqrt{1+x^2}}\,.$$

The first grnm generates a sequence $\{x^s\}_{s=0}^{\infty}$, which converges to x^* from any starting point $x^0 \in R$ with asymptotic quadratic rate. For the starting point $x^0 = 10$ and $\epsilon = 10^{-10}$, we have the following sequence: $\{10;\ 9.005;\ 8.011;\ 7.019;\ 6.029;\ 5.042;\ 4.061;\ 3.090;\ 2.139;\ 1.233;\ 0.456;\ 0.041;\ 3.490\cdot 10^{-5};\ 2.125\cdot 10^{-14}\}$.

Now we consider the second grnm for a convex function $f : \mathbf{R}^n \to \mathbf{R}$, which satisfies (16)–(17) in the neighborhood $S(x^*, \rho)$ of the solution, but generally speaking neither strongly convex nor even smooth on the entire $\mathbf{R}^n$.

First, we introduce the modification of the rnm (31).

We assume $m(x) = 0$ and $M(x) = \infty$ for any given $x \in \mathbf{R}^n$ where the Hessian $\nabla^2 f(x)$ does not exist. For a given $x \in \mathbf{R}^n$, we consider the following matrix

$$A(x) = \begin{cases} \nabla^2 f(x), & 0 < m(x) \le M(x) < \infty; \\ O^{n,n}, & \text{otherwise.} \end{cases} \tag{39}$$

The rnm with step length $t > 0$ is defined by formula

$$\hat{x} := x - t(A(x) + ||\nabla f(x)||I)^{-1} \nabla f(x) = x + tr(x) \tag{40}$$

We will show that the rnm (40) with a special choice of $t > 0$ generates a sequence $\{x^s\}_{s=0}^{\infty}$ that converges to the solution from any starting point for any given convex function $f : \mathbf{R}^n \to \mathbf{R}$.

At the same time, if (16) and (17) are satisfied in the neighborhood $S(x^*, \rho)$, then the rnm (40) converges from any starting point $x \in \mathbf{R}^n$ with asymptotic quadratic rate.

Meanwhile, let us make a few observations about the rnm (40). If the Hessian $\nabla^2 f(x)$ does not exist or $\nabla^2 f(x)$ exists, but $m(x) = 0$, then we cannot find the Newton direction $n(x)$ from (7). On the other hand, from (39) and (40), we obtain

$$\hat{x} := x - t\nabla f(x)(||\nabla f(x)||)^{-1}. \tag{41}$$

if $\nabla f(x)$ exists.

In other words, the rnm (41) turns into the gradient method

$$\hat{x} := x - t(x)\nabla f(x) \tag{42}$$

with step length $t(x) = t||\nabla f(x)||^{-1}$. If the gradient $\nabla f(x)$ satisfies the Lipschitz condition

$$||\nabla f(x) - \nabla f(y)|| \le L||x - y|| \tag{43}$$

then the gradient method (42) with $0 < t(x) < 2L^{-1}$ generates monotone decreasing in value sequence $\{x^s\}$, i.e. $f(x^s) \ge f(x^{s+1})$ and $\lim_{s\to\infty} \nabla f(x^s) = 0$ if $f(x) \ge f(x^*) > -\infty$ (see [7]).

If along with (43) the gradient $\nabla f(x)$ is strongly monotone

$$(\nabla f(x) - \nabla f(y), x - y) \ge m||x - y||^2, \qquad m > 0 \tag{44}$$

then for $0 < t(x) < 2L^{-1}$, the method (42) generates a sequence $\{x^s\}$, which converges to the unique solution with geometric rate. In particular, for $t(x) = L^{-1}$ the following bound holds

$$||x^s - x^*||^2 \le 2m^{-1}q^s(f(x^\circ) - f(x^*)) \tag{45}$$

where $q = 1 - mL^{-1}$ (see [9], pp. 25).

Now, let's assume that $f : \mathbf{R}^n \to \mathbf{R}$ is just convex but not differentiable at $x \in \mathbf{R}^n$, then obviously we cannot find the cnd $n(x)$ from (7) and the classical Newton method cannot be applied at $x \in \mathbf{R}^n$. On the other hand, for any convex function $f : \mathbf{R}^n \to \mathbf{R}$ and any $x \in \mathbf{R}^n$, there exists a subgradient $g(x)$:

$$f(y) - f(x) \geq (g(x), y - x), \qquad \forall y \in \mathbf{R}^n.$$

By setting $\nabla f(x) := g(x)$ and keeping in mind (39) and (40), we obtain

$$\hat{x} := x - tg(x)(||g(x)||)^{-1}, \tag{46}$$

i.e. the rnm (40) turns into the subgradient method.

The properties of the sequence $\{x^s\}$ generated by the subgradient method

$$x^{s+1} = x^s - tg(x^s)||g(x^s)||^{-1} \tag{47}$$

was first established by N. Shor in the early Sixties (see [10] pp 38–39 and references therein).

Let $X_s = \{x : f(x) = f(x^s)\}$, then for a given $\epsilon > 0$ and any $x^* \in X^*$ there is large enough $s_0 > 0$ that the following bound holds (see [10], Theorem 30)

$$\min_{x \in X_{s_0}} ||x - x^*|| \leq \frac{t(1+\epsilon)}{2}.$$

It means that for any given small enough $\epsilon > 0$ there is such a small $t = t(\epsilon)$ and a large number s_0 that the $x^{s_0} \in \{x : f(x^{s_0}) \leq \min_{x \in X^*} f(x) + \epsilon\}$. If (17) holds then x^* is unique and $x^{s_0} \in S(x^*, \rho_0)$ for any $\epsilon > 0$ small enough.

Let $\{t_s\}_{s=0}^{\infty}$ be a positive sequence such that

$$\text{a) } \lim_{s \to \infty} t_s = 0 \qquad \text{b) } \sum_s t_s = \infty. \tag{48}$$

If $f(x)$ is not differentiable, then again the sequence $\{x^s\}_{s=0}^{\infty}$ generated by the rnm (32) turns into the subgradient sequence

$$x^{s+1} = x^s - t_s g(x^s)||g(x^s)||^{-1}. \tag{49}$$

The convergence of the sequence $\{x^s\}_{s=0}^{\infty}$ generated by (49) was established in the Sixties (see [1], [8], [10] and references therein).

Theorem 4. *[10] Let $f : \mathbf{R}^n \to \mathbf{R}$ is convex, the optimal set X^* is bounded and not empty, then the subgradient method (49) with step length (48) generates such a sequence $\{x^s\}_{s=0}^{\infty}$ that either there is s_0 such that $x^{s_0} \in X^*$ or*

$$a) \lim_{s \to \infty} f(x^s) = f(x^*), \text{ and } b) \lim_{s \to \infty} d(x^s, X^*) = 0.$$

There are two main issues with the subgradient method (41).

First, the sequence $\{x^s\}_{s=0}^{\infty}$, generally speaking, is not monotone in value, i.e. for some iterations we can have $f(x^{i+1}) > f(x^i)$.

Second, the subgradient method (49) cannot converge fast because the distance from the current approximation to the solution cannot be less than $t_s > 0$, and t_s according to (48b) converges to zero slow.

Both drawbacks of the subgradient method (49) are less critical, however, for the second grnm, because the method is designed for functions with properties (16) and (17) in the neighborhood of the solution. Therefore the rnm behaves as a gradient (subgradient) method only outside $S(x^*, \rho)$. After reaching the neighborhood $S(x^*, \rho)$ the rnm generates a sequence which due to Theorem 2 converges to the unique minimizer with quadratic rate.

In the rest of the paper, we will be concerned with convex function $f : \mathbf{R}^n \to \mathbf{R}$ for which conditions (16) and (17) are satisfied in the neighborhood $S(x^*, \rho_0)$, but not necessarily on the entire space $\mathbf{R}^n$.

We remind that if $f : \mathbf{R}^n \to R$ is convex and X^* is bounded, then for any given $x^0 \in \mathbf{R}^n$ the level set $\Omega = \{x : f(x) \leq f(x^0)\}$ is bounded.

If the gradient $\nabla f(x)$ does not exist at $x \in \Omega \backslash S(x^*, \rho_0)$ then we set $\nabla f(x) := g(x)$.

Note that from the assumption (17) follows the existence of a small enough $0 < \kappa < 1$ that

$$||\nabla f(x)|| > \kappa, \qquad \forall x \in \Omega \setminus S(x^*, \rho_0). \tag{50}$$

In fact, assuming that (50) is not true, we can find a sequence $\{\kappa_l\}_{l=0}^{\infty}$: $\lim_{l\to\infty} \kappa_l = 0$ and a sequence $\{x^l\}_{l=0}^{\infty} \in \Omega \setminus S(x^*, \rho_0)$ that

$$||\nabla f(x^l)|| \leq \kappa_l \tag{51}$$

The existence of a limit point of $\{x^l\}_{l=0}^{\infty}$ follows from the boundness of Ω. Without restricting the generality we can assume that $\bar{x} = \lim_{l\to\infty} x^l$. Then taking to the limit (51) we obtain

$$||\nabla f(\bar{x})|| = 0, \qquad \bar{x} \notin S(x^*, \rho_0)$$

which is impossible due to the left inequality in (17). Therefore there is $0 < \kappa < 1$ that (50) is true, i.e. if $||\nabla f(x)|| \leq \kappa$ then $x \in S(x^*, \rho_0)$.

We are ready to describe the second grnm.

The second grnm does require neither the existence of $\nabla^2 f(x)$ nor even the existence of $\nabla f(x)$ for all $x \in \mathbf{R}^n$.

We remind that if $\nabla^2 f(x)$ does not exists or $m(x) = 0$ then it follows from (39) and (40) that the regularized Newton direction $r(x) = -\nabla f(x)||\nabla f(x)||^{-1}$ is just the normalized gradient if $\nabla f(x)$ exists or the normalized subgradient $r(x) = -g(x)||g(x)||^{-1}$ if $\nabla f(x)$ does not exist.

There are three critical parameters: $0 < \kappa < 1$, $0 < m_0 \leq m(x)$ and $M(x) \leq M_0 < \infty$, which controls the second grnm. They are usually unknown a priori. Sometimes it is possible to find $0 < \underline{m} \leq m_0$ and $\infty > \overline{M} \geq M_0$, then we set $m_0 := \underline{m}$ and $M_0 := \overline{M}$. If such $0 < \underline{m} < \overline{M} < \infty$ cannot be found a priori, then the second grnm starts with some $0 < \kappa < 1$, $m_0 > 0$ and $M_0 > 0$ and employs

a mechanism which adjusts all three parameters to their appropriate level in the process of solution.

Second Global Regularized Newton Method

Initialization: Given accuracy $\epsilon > 0$, sequence $\{t_s\}_{s=0}^{\infty}$, which satisfies (48), small enough $0 < \kappa < 1$, $0 < \underline{m} < 1$, large enough $\overline{M} > 1$, and a starting point $x^0 \in \mathbf{R}^n$.
Set $\varphi := f(x^0), s := 0, l := 0, \hat{\varphi} := \varphi, \hat{x} := x^0, m_0 := \underline{m}, M_0 := \overline{M}, \kappa_0 := m_0\kappa$.
step 1: $x := \hat{x}$, if $||\nabla f(x)|| \leq \epsilon$, then stop and output $x^* := x$.
step 2: If $||\nabla f(x)|| \leq \kappa_0$ and $m(x) \geq m_0$, $M(x) \leq M_0$, then go to step 7
step 3: Set $A(x) := 0^{n,n}$. If $\nabla f(x)$ does not exist, then set $\nabla f(x) := g(x)$
step 4: $t := t_s$, $s := s + 1$ and find $\hat{x}$ from (40).
step 5: If $f(\hat{x}) < \hat{\varphi}$, then $\hat{\varphi} := f(\hat{x})$ go to step 1.
step 6: Set $x := \hat{x}$, go to step 3.
step 7: Find $r(x)$ from (6), set $t := 1$ and find $\hat{x}$ from (40)
step 8: If $\| \nabla f(\hat{x}) \| > \| \nabla f(x) \|^{1.5}$, then set $t := 0.5 m_0 M_0^{-1}$ and find $\hat{x}$ from (40), $l := l + 1, m_0 := \underline{m} l^{-0.1}, M_0 := \overline{M} l^{0.1}$.
step 9: $\hat{\varphi} := f(\hat{x})$ and go to step 1.

Theorem 5. *Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function and the condition (16) and (17) are satisfied, then there exists s_0 that for*

$$0 < r = (m + 0.5L)m^{-2}||\nabla f(x^{s_0})|| < 1$$

and

$$0 < \rho = m(m + 0.5L)^{-1}r < \rho_0$$

the sequence $\{x^{s_0+s}\}_{s=0}^{\infty} \subset S(x^*, \rho)$ and the following bound

$$||x^{s_0+s} - x^*|| \leq \frac{m}{m + 0.5L} r^{2^s+1} \tag{52}$$

holds for any $s \geq 0$.

[Proof]: We have to show that after finite number of grnm steps we find $x^{s_0} \in S(x^*, \rho_0)$. If $r(x^s) = -\nabla f(x^s)||\nabla f(x^s)||^{-1}$ or $r(x^s) = -g(x^s)||g(x^s)||^{-1}$ is systematically used in (40) for several steps then the existence $x^{s_0} \in S(x^*, \rho_0)$ is a direct consequence of Theorem 4.

Now let's consider the case when after one or few gradient (subgradient) steps the approximation $x^l \notin S(x^*, \rho_0)$ and the step 8 is used.

Note that such step can happen only after the strong monotonicity of the record function value is restored, i.e $\varphi_l < \varphi_{l-1}$.

Let's estimate the lower bound for the $f(x)$ reduction at such a step.

We remind that step 8 with starting point $x = x^l$ is used when $\nabla^2 f(x^l)$ exists, $m(x^l) \geq m_0$, $M(x^l) \leq M_0$ and $||\nabla f(x^{l+1})|| > ||\nabla f(x^l)||^{1.5}$

From (6), we obtain

$$H(x^l)r(x^l) = -\nabla f(x^l)$$

Therefore

$$\begin{aligned}(\nabla f(x^l), r(x^l)) &= -(H(x^l)r(x^l), r(x^l)) \\ &= -(\nabla^2 f(x^l)r(x^l), r(x^l)) - ||\nabla f(x^l)||||r(x^l)||^2 \\ &\leq -m_0||r(x^l)||^2 \end{aligned} \tag{53}$$

For a twice differentiable at the point x^l function we have (see [9], pp. 7)

$$|f(x^l + tr(x^l)) - (f(x^l) + t(\nabla f(x^l), r(x^l)) + \frac{1}{2}t^2(\nabla^2 f(x^l)r(x^l), r(x^l))| \leq o(||tr(x^l)||^2)$$

Keeping in mind the conditions from step 2 and (6), we obtain

$$||r(x^l)|| \leq ||(H(x^l))^{-1}||||\nabla f(x^l)|| \leq m_0^{-1}\kappa_0 \leq \kappa$$

Therefore for $0 < t \leq 1$ and small enough $\kappa > 0$ we obtain $o(||tr(x^l)||^2) = \alpha_l||tr(x^l)||^2$ and $\alpha_l \to 0$.

Using (53) and bounds for $m(x^l)$ and $M(x^l)$ we obtain

$$\begin{aligned}f(x^l + tr(x^l)) &\leq f(x^l) + t(\nabla f(x^l), r(x)) + \frac{t^2}{2}(\nabla^2 f(x^l)r(x^l), r(x^l))| + o(t^2||r(x^l)||^2) \\ &\leq f(x^l) - tm_0||r(x^l)||^2 + t^2 M_0||r(x^l)||^2 \\ &= f(x^l) - t(m_0 - tM_0)||r(x^l)||^2\end{aligned}$$

Hence, for $t = 0.5m_0M_0^{-1}$ we have

$$\begin{aligned}f(x^{l+1}) = f(x^l + tr(x^l)) &\leq f(x^l) - 0.25m_0^2M_0^{-1}||r(x^l)||^2 \\ &= f(x^l) - \sigma l^{-0.3}||r(x^l)||^2 \end{aligned} \tag{54}$$

where $\sigma = 0.25\underline{m}^2\overline{M}^{-1}$.

Summing up (54) from $l = 0$ to $l = \infty$, we obtain

$$\sigma\sum_{l=0}^{\infty} l^{-0.3}||r(x^l)||^2 < f(x^0) - f(x^*).$$

Therefore there is $l = s_0$ such that $||r(x^{s_0})|| \leq l^{-0.35}$.

From (9) and $||\nabla f(x^{s_0})|| \leq \kappa_0 = m_0\kappa, m_0 = \underline{m}s_0^{-0.1}, M_0 = \overline{M}s_0^{0.1}$, we obtain

$$\begin{aligned}||\nabla f(x^{s_0})|| &\leq ||H(x^{s_0})||||r(x^{s_0})|| \\ &\leq (||\nabla^2 f(x^{s_0})|| + ||\nabla f(x^{s_0})||)||r(x^{s_0})|| \\ &\leq (M_0 + \kappa_0)||r(x^{s_0})|| \\ &\leq 2\overline{M}s_0^{-0.25}\end{aligned}$$

Hence, for a given small enough $0 < \kappa < 1$, there is $l = s_0 > 0$ that

$$||\nabla f(x^{s_0})|| \leq 2\overline{M} s_0^{-0.25} \leq \kappa \tag{55}$$

It follows from (50) and (55) that $x^{s_0} \in S(x^*, \rho_0)$.

The bound (52) is a direct consequence of Theorem 2. □

Example 2.

$$f_1(x) = \begin{cases} -3x - 2, & -\infty < x \leq -1; \\ \frac{1}{2}\left(x^2 + x^4\right), & -1 \leq x \leq 1; \\ 3x - 2, & 1 \leq x < \infty \, . \end{cases}$$

$$f_2(x) = \max\left\{\frac{16}{3}x - 8, -\frac{16}{3}x - 8\right\}$$

$$f(x) = \max\{f_1(x), f_2(x)\}$$

Using initialization $\epsilon = 10^{-15}$, $t_s = s^{-1}$, $\kappa = 0.1$, $\underline{m} = 0.9$, $\overline{M} = 7$ and starting point $x^0 = 3.0$, the second grmn generates the following sequence: $\{3.0; 2.0; 1.5; 1.17; 0.92; 0.63; 0.40; 0.21; 0.064; 0.0063; 6.6 \times 10^{-5}; 1.2 \times 10^{-8}; 8.3 \times 10^{-16}; \cdots\}$.

5. Concluding remarks.

The most costly operation of the grnm is solving the system (6) to find $r(x)$. The Cholesky factorization of $H(x)$ is an efficient tool for solving the system (6). It is worth to mention that by using Cholesky factorization along with $r(x)$, one finds $m(x)$ and $M(x)$ practically with very little extra numerical computation.

It follows from Theorem 2 that the sooner the approximation ends up in $S(x^*, \rho)$, the better it is.

The size $S(x^*, \rho)$ depends on $m > 0$, L and $0 < r < 1$. For a smaller r it takes longer to get into $S(x^*, \rho)$, but when the approximation is in $S(x^*, \rho)$ the convergence is faster.

The grnms substantially enlarge the class of convex functions for which Newton method can be applied. Moreover, the grnms retain the most important property of the Newton method - its quadratic rate of convergence in the neighborhood of the solution.

We believe, however, that there is room for improvement on both theoretical and numerical sides. On the theoretical side, finding the optimal step length for the first grnm might help to improve the complexity bound (38). On the numerical side, we have to conduct extensive numerical experiments to better understand the numerical efficiency of grnms.

Acknowledgements. The author is grateful to Yu. Nesterov and the anonymous referee for their valuable comments, which helped to improve the paper.

The research was partially supported by NSF Grant CCF-0324999.

References

1. Yu. M. Ermoliev, "Methods for solving nonlinear extremal problems", Kibernetika **4,** (1966) 1-17 (in Russian).
2. L. V. Kantorovich, "Functional analysis and applied mathematics", Uspekhi Mat. Nauk **3,** (1948) 89-185 (in Russian).
3. K. Levenberg, "A method for the solution of certain nonlinear problems in least squares", Quart. Appl. Math. **2,** (1944) 164-168.
4. D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters", J. Soc. Indust. Appl. Math. **11,** (1963), 431-441.
5. Yu. Nesterov and A. Nemirovsky, *Interior Point Polynomial Algorithms in Convex Programming* (SIAM, Philadelphia, 1994)
6. Yu. Nesterov and B. T. Polyak, "Cubic Regularization of Newton Method and its Global Performance", Mathematical Programming, DOI 10.1007/s 10107-006-0706-8, published online 04/25/06.
7. B. T. Polyak, "Gradient Methods for the Minimization of Functionals", USSR Comp. Math. and Math. Phys. **3,** No. 4, (1963), 643-653.
8. B. T. Polyak, "A General Method of Solving Extremum Problems", Soviet Math. Dokl. **3,** (1967), 593-597.
9. B. T. Polyak, *Introduction to Optimization* (New York, Optimization Software, 1987)
10. N. Z. Shor, *Nondifferentiable Optimization and Polynomial Problems* (Kluwer Academic Publishers, Boston, 1998)
11. S. Smale, "Newton's method estimates from data at one point", in The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics, R. E. Ewing, K.I. Gross, C. F. Martin, eds., Springer-Verlag, New York, (1986), 185-196.