COMPLEXITY OF THE REGULARIZED NEWTON METHOD

ROMAN A. POLYAK

Abstract. Newton's method for finding unconstrained minimizer of strictly convex functions, generally speaking, does not converge from any starting point.

We introduce and study the damped regularized Newton's method (DRNM). It converges globally for any strictly convex function, which has a minimizer in $\mathbb{R}^n$.

Locally DRNM converges with quadratic rate. We characterize the neighborhood of the minimizer, where the quadratic rate occurs. Based on it we estimate the number of DRNM's steps required for finding an ε- approximation for the minimizer.

1. Introduction

Newton's method, which has been introduced almost 350 years ago, is still one of the basic tools in numerical analysis, variational and control problems, optimization both constrained and unconstrained, just to mention a few.

It has been used not only as a numerical tool, but also as a powerful instrument for proving existence and uniqueness results.

In particular, Newton-Kantorovich's method plays a critical role in the classical KAM theory by Kolmogorov, Arnold and Mozer (see [1]). Another example is the proof of Lusternik's theorem on tangent spaces (see [3], [9]).

Newton's method was the main instrument in the interior point methods (IPMs), which preoccupied the field of optimization for a long time.

Yu. Nesterov and A. Nemirovski shown that a special damped Newton's method is particularly efficient for minimization self - concordant (SC) functions (see [6], [7]).

They shown that from any starting point a special damped Newton's step reduces the SC function value by a constant, which depends only on the Newton's decrement. The decrement converges to zero.

By the time it gets small enough the damped Newton's method practically turns into Newton's method and generates a sequence, which converges in value with quadratic rate.

They characterized the size of the minimizer's neighborhood, where quadratic rate occurs. It allows establishing the complexity of the special

2010 *Mathematics Subject Classification.* 65K10 80M50 90C26.

Key words and phrases. Regularized Newton's Method, Newton's Decrement, Newtoin's Area, Global Convergense, Quadratic Convergence Rate.

damped Newton's method for SC function, that is to find the upper bound for the number of damped Newton's step required for finding an ε- approximation for the minimizer.

For strictly convex functions, which are not self-concordant, such results, to the best of our knowledge, are unknown.

The purpose of the paper is to introduce and establish complexity bounds of the damped Newton's method (DNM) and DRNM for minimization of twice continuously differentiable and strictly convex $f : \mathbb{R}^n \to \mathbb{R}$.

First, we characterize the Newton's areas for DNM and DRNM. In other words, we estimate the minimizer's neighborhoods, where DNM and DRNM converges with quadratic rate.

Then we estimate the number of steps needed for DNM's or DRNM's to enter the correspondent Newton's areas.

The key ingredients of our analysis are the Newton's and the regularized Newton's decrements.

On the one hand, the decrements provide the upper bound for the distance from the current approximation to the minimizer. Therefore they have been used in the stopping criteria.

On the other hand, they provide a lower bound for the function reduction at each step at any point, which does not belong to the Newton's or to the regularized Newton's area.

These bounds were used to estimate the number of DNM or DRNM steps needed to get into the correspondent Newton's areas.

2. Newton's Method

We start with the classical Newton's method for finding a root of a nonlinear equation

$$f(t) = 0,$$

where $f : \mathbb{R} \to \mathbb{R}$ has a smooth derivative f'.

Let us consider $t_0 \in \mathbb{R}$ and the linear approximation

$$\widetilde{f}(t) = f(t_0) + f'(t_0)(t - t_0) = f(t_0) + f'(t_0)\Delta t$$

of f at t_0, assuming that $f'(t_0) \neq 0$.

By replacing f with its linear approximation we obtain the following equation

$$f(t_0) + f'(t_0)\Delta t = 0$$

for the Newton's step Δt.

The next approximation is given by formula

$$t = t_0 + \Delta t = t_0 - (f'(t_0))^{-1} f(t_0). \tag{2.1}$$

By reiterating (2.1) we obtain Newton's method

$$t_{s+1} = t_s - (f'(t_s))^{-1} f(t_s) \tag{2.2}$$

for finding a root of a nonlinear equation $f(t) = 0$.

Let t^* be the root, that is $f(t^*) = 0$. Also we assume $f^{'}(t^*) \neq 0$ and $f \in C^2$. We consider the expansion of f at t_s with the Lagrange remainder

$$0 = f(t^*) = f(t_s) + f^{'}(t_s)(t^* - t_s) + \frac{1}{2}f^{''}(\hat{t}_s)(t^* - t_s)^2, \tag{2.3}$$

where $\hat{t}_s \in [t_s, t^*]$. For t_s close to t^* we have $f^{'}(t_s) \neq 0$, therefore from (2.3) follows

$$t^* - t_s + \frac{f(t_s)}{f'(t_s)} = -\frac{1}{2}\frac{f^{''}(\hat{t}_s)}{f'(t_s)}(t^* - t_s)^2.$$

Using (2.2) we get

$$|t^* - t_{s+1}| = \frac{1}{2}\frac{|f^{''}(\hat{t}_s)|}{|f'(t_s)|}|t^* - t_s|^2. \tag{2.4}$$

If $\Delta_s = |t^* - t_s|$ is small, then there exist $a > 0$ and $b > 0$ independent on t_s that $|f^{''}(\hat{t}_s)| \leq a$ and $|f^{'}(t_s)| > b$. Therefore, from (2.4) follows

$$\Delta_{s+1} \leq c\Delta_s^2, \tag{2.5}$$

where $c = 0.5ab^{-1}$.

This is the key characteristic of Newton's method, which makes the method so important even 350 years after it was originally introduced.

Newton's method has a natural extension for a nonlinear system of equations

$$g(x) = 0, \tag{2.6}$$

where $g : \mathbb{R}^n \to \mathbb{R}^n$ is a vector-function with a smooth Jacobian $J(g) = \nabla g : \mathbb{R}^n \to \mathbb{R}^n$. The linear approximation of g at x_0 is given by

$$\widetilde{g}(x) = g(x_0) + \nabla g(x_0)(x - x_0). \tag{2.7}$$

We replace g in (2.6) by its linear approximation (2.7). The Newton's step Δx one finds by solving the following linear system :

$$g(x_0) + \nabla g(x_0)\Delta x = 0.$$

Assuming $\det \nabla g(x_0) \neq 0$ we obtain

$$\Delta x = -(\nabla g(x_0))^{-1} g(x_0).$$

The new approximation is given by the following formula:

$$x = x_0 - (\nabla g(x_0))^{-1} g(x_0). \tag{2.8}$$

By reiterating (2.8) we obtain Newton's method

$$x_{s+1} = x_s - (\nabla g(x_s))^{-1} g(x_s) \tag{2.9}$$

for solving a nonlinear system of equations (2.6).

Newton's method for minimization of $f : \mathbb{R}^n \to \mathbb{R}$ follows directly from (2.9) if instead of unconstrained minimization problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in \mathbb{R}^n \end{aligned} \tag{2.10}$$

we consider the nonlinear system

$$\nabla f(x) = 0, \tag{2.11}$$

which is the necessary and sufficient condition for x^* to be the minimizer in (2.10) in case of convex f.

Vector

$$n(x) = -(\nabla^2 f(x))^{-1} \nabla f(x) \tag{2.12}$$

defines the Newton's direction at $x \in \mathbb{R}^n$.

Application of Newton's method (2.9) to the system (2.11) leads to the Newton's method

$$x_{s+1} = x_s - (\nabla^2 f(x_s))^{-1} \nabla f(x_s) = x_s + n(x_s) \tag{2.13}$$

for solving (2.10).
Method (2.13) has another interpretation. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with a positive definite Hessian $\nabla^2 f$.

The quadratic approximation of f at x_0 is given by the formula

$$\widetilde{f}(x) = f(x_0) + (\nabla f(x_0), x - x_0) + \frac{1}{2}(\nabla^2 f(x_0)(x - x_0), x - x_0).$$

Instead of solving (2.10) let us find

$$\bar{x} = \text{argmin}\{\widetilde{f}(x) \, : x \in \mathbb{R}^n\},$$

which is equivalent to solving the following linear system

$$\nabla^2 f(x_0) \Delta x = -\nabla f(x_0)$$

for $\Delta x = x - x_0$.
We obtain

$$\Delta x = n(x_0),$$

so for the next approximation we have

$$\bar{x} = x_0 - (\nabla^2 f(x_0))^{-1} \nabla f(x_0) = x_0 + n(x_0). \tag{2.14}$$

By reiterating (2.14) we obtain Newton's method (2.13) for solving (2.10).

The local quadratic convergence of both (2.9) and (2.13) is well known (see [2], [4], [7], [8] and references therein).

Away from the neighborhood of x^*, however, both Newton's methods (2.9) and (2.13) can either oscillate or diverge.

Example 2.1. Consider

$$g(t) = \begin{cases} -(t-1)^2 + 1, & t \geq 0, \\ \quad (t+1)^2 - 1, & t < 0. \end{cases}$$

The function g together with g' is continuous on $(-\infty, \infty)$. Newton's method (2.2) converges to the root $t^* = 0$ from any starting point t: $|t| < \frac{2}{3}$, oscillates between $t_s = -\frac{2}{3}$ and $t_{s+1} = \frac{2}{3}$, $s = 1, 2, \ldots$ and diverges for any t: $|t| > \frac{2}{3}$.

Example 2.2. For $f(t)=\sqrt{1+t^2}$ we have

$$f(t^*)=f(0)=\min\{f(t)\,:\,-\infty<t<\infty\}.$$

For the first and second derivative we have

$$f'(t)=t(1+t^2)^{-\frac{1}{2}},\ \ f''(t)=(1+t^2)^{-\frac{3}{2}}.$$

Therefore Newton's method (2.13) is given by the following formula

$$t_{s+1}=t_s-(1+t_s^2)^{\frac{3}{2}}t_s(1+t_s^2)^{-\frac{1}{2}}=-t_s^3. \tag{2.15}$$

It follows from (2.15) that Newton's method converges from any $t_0\in(-1,1)$ oscillates between $t_s=-1$ and $t_{s+1}=1$, $s=1,2,...$ and diverges from any $t_0\notin[-1,1]$. It also follows from (2.15) that Newton's method converges from any starting point $t_0\in(-1,1)$ with the cubic rate, however, in both examples the convergence area is negligibly smaller than the area where Newton's method diverges. Note that f is strictly convex in $\mathbb{R}$ and strongly convex in the neighborhood of $t^*=0$.

Therefore there are three important issues associated with the Newton's method for unconstrained convex optimization.
First, to characterize the neighborhood of the solution, where Newton's method converges with quadratic rate.
Second, to find such modification of Newton's method that generates convergent sequence from any starting point and retains quadratic convergence rate in the neighborhood of the solution.
Third, to estimate the computational complexity of a globally convergent Newton's and regularized Newton's methods in terms of the total number of steps required for finding an ε-approximation for x^*.

3. Local Quadratic Convergence of Newton's Method

We consider a class of convex functions $f:\mathbb{R}^n\to\mathbb{R}$, that are strongly convex at x^*, that is

$$\nabla^2 f(x^*)\succeq mI, \tag{3.1}$$

$m>0$ and their Hessian satisfy Lipschitz condition in the neighborhood of x^*. In other words there is $\delta>0$, a ball $B(x^*,\delta)=\{x\in\mathbb{R}^n,\|x-x^*\|\le\delta\}$ and $M>0$ such that for any x and $y\in B(x^*,\delta)$ we have

$$\|\nabla^2 f(x)-\nabla^2 f(y)\|\le M\|x-y\|. \tag{3.2}$$

The following Theorem characterize the neighborhood of x^*, where Newton's method converges with quadratic rate.

There are several ways to proof this fundamental result(see, for example, [2], [4], [7], [8] and references therein). In the following Theorem, which we provide for completeness, the Newton's area is characterized explicitly through the convexity constant $m>0$ and Lipschitz constant $M>0$ (see [7]). We will use these technique later to characterize the regularized Newton's area.

Theorem 3.1. *If for $0 < m < M$ conditions* (3.1) *and* (3.2) *are satisfied, then for $\delta = \frac{2m}{3M}$ and any given $x_0 \in B(x^*, \delta)$ the entire sequence $\{x_s\}_{s=0}^{\infty}$ generated by (2.13) belongs $B(x^*, \delta)$ and the following bound holds:*

$$\|x_{s+1} - x^*\| \le \frac{M}{2(m - M\|x_s - x^*\|)}\|x_s - x^*\|^2,\ s \ge 1. \tag{3.3}$$

Proof. From (2.13) and $\nabla f(x^*) = 0$ follows

$$x_{s+1} - x^* = x_s - x^* - [\nabla^2 f(x_s)]^{-1}\nabla f(x_s) =$$
$$= x_s - x^* - (\nabla^2 f(x_s))^{-1}(\nabla f(x_s) - \nabla f(x^*)) =$$
$$= [\nabla^2 f(x_s)]^{-1}[\nabla^2 f(x_s)(x_s - x^*) - (\nabla f(x_s) - \nabla f(x^*))]. \tag{3.4}$$

Then we have

$$\nabla f(x_s) - \nabla f(x^*) = \int_0^1 \nabla^2 f(x^* + \tau(x_s - x^*))(x_s - x^*)d\tau.$$

From (3.4) we obtain

$$x_{s+1} - x^* = [\nabla^2 f(x_s)]^{-1} H_s(x_s - x^*), \tag{3.5}$$

where

$$H_s = \int_0^1 [\nabla^2 f(x_s) - \nabla^2 f(x^* + \tau(x_s - x^*))]d\tau.$$

Let $\Delta_s = \|x_s - x^*\|$, then using (3.2) we get

$$\|H_s\| = \|\int_0^1 [\nabla^2 f(x_s) - \nabla^2 f(x^* + \tau(x_s - x^*))]d\tau\|$$
$$\le \int_0^1 \|[\nabla^2 f(x_s) - \nabla^2 f(x^* + \tau(x_s - x^*))\|d\tau \le$$
$$\le \int_0^1 M\|x_s - x^* - \tau(x_s - x^*)\|d\tau \le$$
$$\le \int_0^1 M(1 - \tau)\|x_s - x^*\|d\tau = \frac{M}{2}\Delta_s.$$

Therefore from (3.5) and the latter bound we have

$$\Delta_{s+1} \le \|(\nabla^2 f(x_s))^{-1}\|\|H_s\|\|x_s - x^*\| \le$$
$$\frac{M}{2}\|(\nabla^2 f(x_s))^{-1}\|\Delta_s^2. \tag{3.6}$$

From (3.2) follows

$$\|\nabla^2 f(x_s) - \nabla^2 f(x^*)\| \le M\|x_s - x^*\| = M\Delta_s,$$

therefore

$$\nabla^2 f(x^*) + M\Delta_s I \succeq \nabla^2 f(x_s) \succeq \nabla^2 f(x^*) - M\Delta_s I.$$

From (3.1) follows

$$\nabla^2 f(x_s) \succeq \nabla^2 f(x^*) - M\Delta_s I \succeq (m - M\Delta_s)I.$$

Hence, for any $\Delta_s < mM^{-1}$ the matrix $\nabla^2 f(x_s)$ is positive definite, therefore the inverse $(\nabla^2 f(x^s))^{-1}$ exists and the following bound holds

$$\|(\nabla^2 f(x^s))^{-1}\| \leq \frac{1}{m - M\Delta_s}.$$

From (3.6) and the latter bound follows

$$\Delta_{s+1} \leq \frac{M}{2(m - M\Delta_s)}\Delta_s^2. \tag{3.7}$$

From (3.7) for $\Delta_s < \frac{2m}{3M}$ follows $\Delta_{s+1} < \Delta_s$, which means that for $\delta = \frac{2m}{3M}$ and any $x_0 \in B(x^*, \delta)$ the entire sequence $\{x_s\}_{s=0}^{\infty}$ belongs to $B(x^*, \delta)$ and converges to x^* with the quadratic rate (3.7).

The proof is completed □

The neighborhood $B(x^*, \delta)$ with $\delta = \frac{2m}{3M}$ is called Newton's area.

In the following section we consider a new version of the damped Newton's method, which converges from any starting point and at the same time retains quadratic convergence rate in the Newton's area.

4. Damped Newton's Method

To make Newton's method practical we have to guarantee convergence from any starting point. To this end the step length $t > 0$ is attached to the Newton's direction $n(x)$, that is

$$\hat{x} = x + tn(x) = x - t(\nabla^2 f(x))^{-1}\nabla f(x). \tag{4.1}$$

The step length $t > 0$ has to be adjusted to guarantee a "substantial reduction" of f at each $x \notin B(x^*, \delta)$ and $t = 1$, when $x \in B(x^*, \delta)$.
Method (4.1) is called the damped Newton's Method (DNM)(see, for example, [2], [7], [9])

The following function $\lambda : \mathbb{R}^n \to \mathbb{R}_+$:

$$\lambda(x) = ((\nabla^2 f(x))^{-1}\nabla f(x), \nabla f(x))^{0.5} = [-(\nabla f(x), n(x))]^{0.5}, \tag{4.2}$$

which is called the Newton's decrement of f at $x \in \mathbb{R}^n$, will play an important role later.

At this point we assume that $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex and its Hessian $\nabla^2 f$ is Lipschitz continuous, that is, there exist $\infty > M > m > 0$ that

$$\nabla^2 f(x) \succeq mI \tag{4.3}$$

and

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq M\|x - y\| \tag{4.4}$$

are satisfied for any x and y from $\mathbb{R}^n$.

Let $x_0 \in \mathbb{R}^n$ be a starting point.

Due to (4.3) the sublevel set $\mathcal{L}_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded for any given $x_0 \in \mathbb{R}^n$. Therefore from (4.4) follows existence $L > 0$ that

$$\|\nabla^2 f(x)\| \leq L \tag{4.5}$$

is taking place.

We also assume that $\varepsilon > 0$ is small enough, in particular,

$$0 < \varepsilon < m^2 L^{-1} \tag{4.6}$$

holds.

We are ready to describe our version of DNM.

Let $x_0 \in \mathbb{R}^n$ be a starting point and $0 < \varepsilon < \delta$ be the required accuracy. Set $x := x_0$

1. find Newton's direction $n(x)$;
2. if the following inequality

$$f(x + n(x)) \leq f(x) + 0.5(\nabla f(x), n(x)) \tag{4.7}$$

   holds, then set $t(x) := 1$, otherwise set $t(x) = m(2L)^{-1}$;
3. set $x := x + t(x)n(x)$;
4. if $\lambda(x) \leq \varepsilon^{1.5}$, then $x^* := x$, otherwise go 1.

The following Theorem proves global convergence of the DNM 1.-4. and establishes the upper bound for the total number of DNM steps require for finding ε-approximation for x^*.

5. Global Convergence of the DNM and its Complexity

Theorem 5.1. *If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and conditions (4.3) and (4.4) are satisfied, then for $\delta = \frac{2}{3}\frac{m}{M}$ it takes*

$$N_0 = 9\frac{L^2 M^2}{m^5}(f(x_0) - f(x^*)). \tag{5.1}$$

DNM steps to find $x \in B(x^*, \delta)$ by using DNM.

Proof. From (4.5) follows

$$\nabla^2 f(x) \preceq LI. \tag{5.2}$$

On other hand, from (4.3) follows the existence of the inverse $(\nabla^2 f(x))^{-1}$. Therefore from (5.2) follows

$$(\nabla^2 f(x))^{-1} \succeq L^{-1} I. \tag{5.3}$$

From (4.2) and (5.3) we obtain the following lower bound for the Newton's decrement

$$\begin{aligned} \lambda(x) &= (\nabla^2 f(x)^{-1} \nabla f(x), \nabla f(x))^{0.5} \geq \\ &\geq (L^{-1}\|\nabla f(x)\|^2)^{0.5} = L^{-0.5}\|\nabla f(x)\|. \end{aligned} \tag{5.4}$$

From (4.3) we have

$$\|\nabla f(x)\|\|x - x^*\| \geq (\nabla f(x) - \nabla f(x^*), x - x^*) \geq m\|x - x^*\|^2$$

or

$$\|\nabla f(x)\| \geq m\|x - x^*\|. \tag{5.5}$$

From (5.4) and (5.5) we obtain

$$\lambda(x) \geq L^{-0.5} m \|x - x^*\|. \tag{5.6}$$

From (4.6) and the stopping criteria 4. follows

$$(m^2 L^{-1})^{0.5} \varepsilon \geq \varepsilon^{1.5} \geq \lambda(x) \geq m L^{-0.5} \|x - x^*\|,$$

or $\|x - x^*\| \leq \varepsilon$, which justifies the stopping criteria 4.

On the other hand, Newton's decrement defines the lower bound for the function reduction at each step.

In fact, for Newton's directional derivative from (2.12), (4.2) and (4.3) follows

$$\varphi'(0) = \frac{df(x + tn(x))}{dt}|_{t=0} = (\nabla f(x), n(x)) =$$

$$-(\nabla^2 f(x) n(x), n(x)) \leq -m\|n(x)\|^2. \tag{5.7}$$

Due to the strong convexity of $\varphi(t) = f(x + tn(x))$ the derivative $\varphi'(t) = (\nabla f(x + tn(x)), n(x))$ is monotone increasing in $t > 0$, so there is $t(x) > 0$ such that

$$0 > (\nabla f(x + t(x)n(x)), n(x)) \geq \frac{1}{2}(\nabla f(x), n(x)), \tag{5.8}$$

otherwise $(\nabla f(x + tn(x)), n(x)) < \frac{1}{2}(\nabla f(x), n(x)) \leq -\frac{1}{2} m\|n(x)\|^2, t > 0$ and $\inf f(x) = -\infty$, which is impossible for a strongly convex function f.

It follows from (5.7), (5.8) and monotonicity of $\varphi'(t)$ that for any $t \in [0, t(x)]$ we have

$$\frac{df(x + tn(x))}{dt} = (\nabla f(x + tn(x)), n(x)) \leq \frac{1}{2}(\nabla f(x), n(x)).$$

Therefore

$$f(x + t(x)n(x)) \leq f(x) + \frac{1}{2} t(x)(\nabla f(x), n(x)).$$

Keeping in mind (4.2) we obtain

$$f(x) - f(x + t(x)n(x)) \geq \frac{1}{2} t(x) \lambda^2(x). \tag{5.9}$$

Combining (5.7) and (5.8) we obtain

$$(\nabla f(x + t(x)n(x)) - \nabla f(x), n(x)) \geq \frac{m}{2}\|n(x)\|^2.$$

Therefore, there is $0 < \theta(x) < 1$ such that

$$t(x)(\nabla^2 f(x + \theta(x)t(x)n(x))n(x), n(x)) = t(x)(\nabla^2 f(\cdot)n(x), n(x)) \geq \frac{m}{2}\|n(x)\|^2,$$

or

$$t(x)\|\nabla^2 f(\cdot)\|\|n(x)\|^2 \geq \frac{m}{2}\|n(x)\|^2.$$

From (4.3) follows

$$t(x) \geq \frac{m}{2L}, \tag{5.10}$$

which justifies the choice of step length $t(x)$ in the DNM 1.-4..

Hence, from (5.9) and (5.10) we obtain the following lower bound for the function reduction per step

$$\Delta f(x) = f(x) - f(x + t(x)n(x)) \geq \frac{m}{4L}\lambda^2(x), \tag{5.11}$$

which together with the lower bound (5.6) for the Newton's decrement $\lambda(x)$ leads to

$$\Delta f(x) = f(x) - f(x + t(x)n(x)) \geq \frac{m^3}{4L^2}\|x - x^*\|^2. \tag{5.12}$$

It means that for any $x \notin B(x^*, \delta)$ the function reduction at each step is proportional to the square of the distance between current approximation x and the solution x^*.

In other words, "far from" the solution Newton's step produces a "substantial" reduction of the function value similar to one of the gradient method.

For $x \notin B(x^*, \delta)$ we have $\|x - x^*\| \geq \frac{2m}{3M}$, therefore from (5.12) we obtain $\Delta f(x) \geq \frac{1}{9}\frac{m^5}{L^2M^2}$. So it takes at most

$$N_0 = 9\frac{L^2M^2}{m^5}(f(x_0) - f(x^*))$$

Newton's steps to obtain $x \in B(x^*, \delta)$ from a given starting point $x_0 \in \mathbb{R}^n$. The proof is completed. □

From Theorem 3.1 follows that $O(\ln\ln\varepsilon^{-1})$ steps needed to find an ε-approximation to x^* from any $x \in B(x^*, \delta)$, where $0 < \varepsilon < \delta$ is the required accuracy. Therefore the total number of Newton's steps required for finding an ε-approximation to the optimal solution x^* from a starting point $x_0 \in \mathbb{R}^n$ is

$$N = N_0 + O(\ln\ln\varepsilon^{-1}).$$

The bound (5.1) is similar to (9.40) from [2], but the proof is based on our version of DNM and the explicit characterization of the Newton's area. It allows to extend the proof for the regularized Newton's method [10].

The DNM requires an a priori knowledge of two parameters m and L or their corresponding lower and upper bounds.

The following version of DNM is free from this requirement. To adjust the step length $t > 0$ we use the backtracking line search.

The inequality

$$f(x + tn(x)) \leq f(x) + \alpha t(\nabla f(x), n(x)) \tag{5.13}$$

with $0 < \alpha \leq 0.5$ is called the Armijo condition.

Let $0 < \rho < 1$, the backtracking line search consist of the following steps.

1. For $t > 0$ check (5.13). If (5.13) holds go to 2. If not set $t := t\rho$ and repeat it until (5.13) holds, then go to 2.
2. set $t(x) := t$, $x := x + t(x)n(x)$

We are ready to describe another version of DNM, which does not requires an a priori knowledge of the parameters m and L or their lower and upper bounds.

Let $x_0 \in \mathbb{R}^n$ be a starting point and $0 < \varepsilon << \delta$ be the required accuracy.

1. Compute Newton's direction

$$n(x) = -(\nabla^2 f(x))^{-1} \nabla f(x); \tag{5.14}$$

2. set $t := 1$, use the backtracking line search until

$$f(x + tn(x)) \leq f(x) + 0.5t(\nabla f(x), n(x));$$

3. set $t(x) := t$, $x := x + t(x)n(x)$;
4. if $\lambda(x) \leq \varepsilon^{1.5}$ then $x^* := x$ otherwise go 1.

The complexity of the DNM with backtracking line search can be established using arguments similar to those in Theorem 5.1

Unfortunately, in the absence of strong convexity of $f : \mathbb{R}^n \to \mathbb{R}$ Newton's method might not converge from any starting point.

In case of Example 2.2 Newton's method does not converge from any $t \notin (-1, 1)$ in spite of $f(t) = \sqrt{1 + t^2}$ being strongly convex and smooth enough in the neighborhood of $t^* = 0$.

In the following section we consider the Regularized Newton's Method (RNM)(see [10]), which eliminates the basic drawback of the Classical Newton's Method. It generates a converging sequence from any starting point $x_0 \in \mathbb{R}^n$ and retains quadratic convergence rate in the regularized Newton's area, which we will characterize later.

6. Regularized Newton's Methods

Let $f \in C^2$ be a convex function in $\mathbb{R}^n$. We assume that the optimal set $X^* = \text{Argmin}\{f(x) : x \in \mathbb{R}^n\}$ is not empty and bounded.

The corresponding regularized at the point $x \in \mathbb{R}^n$ function $F_x : \mathbb{R}^n \to \mathbb{R}$ is defined by the following formula

$$F_x(y) = f(y) + \frac{1}{2} \parallel \nabla f(x) \parallel \parallel y - x \parallel^2 . \tag{6.1}$$

For any $x \notin X^*$ we have $||\nabla f(x)|| > 0$, therefore for any convex function $f : \mathbb{R}^n \to \mathbb{R}$ the regularized function F_x is strongly convex in y for any $x \notin X^*$. If f is strongly convex at x^*, then the regularized function F_x is strongly convex in $\mathbb{R}^n$. The following properties of F_x are direct consequences of the definition (6.1).

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 identical matrix in $\mathbb{R}^n$.

For any $x \notin X^*$, the inverse $H^{-1}(x)$ exists for any convex $f \in C^2$. Therefore the regularized Newton's step

$$\hat{x} = x - (H(x))^{-1} \nabla f(x) \tag{6.2}$$

can be performed for any convex $f \in C^2$ from any starting point $x \notin X^*$.

We start by showing that the regularization (6.1) improves the "quality" of the Newton's direction as well the condition number of the Hessian $\nabla^2 f(x)$ at any $x \in \mathbb{R}^n$ that $x \notin X^*$.

We assume at this point that for any given $x \in \mathbb{R}^n$ there exist $0 \leq m(x) < M(x) < \infty$ such that

$$m(x)||y||^2 \leq (\nabla^2 f(x)y, y) \leq M(x)||y||^2 \tag{6.3}$$

holds for any $y \in \mathbb{R}^n$.

The condition number of the Hessian $\nabla^2 f$ at $x \in \mathbb{R}^n$ is

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

Along with the regularized Newton's step (6.2), we consider the classical Newton's step

$$\hat{x} = x - (\nabla^2 f(x))^{-1}\nabla f(x). \tag{6.4}$$

The regularized Newton's direction (RND) $r(x)$ is defined by the system

$$H(x)r(x) = -\nabla f(x). \tag{6.5}$$

The "quality" of any direction d at $x \in \mathbb{R}^n$ is define by the following number

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

For the steepest descent direction $d(x) = -\nabla f(x) \parallel \nabla f(x) \parallel^{-1}$ we have the best local descent direction and $q(d(x)) = 1$. The "quality" of the classical Newton's direction is defined by the following number

$$q(n(x)) = -\frac{(\nabla f(x), n(x))}{||\nabla f(x)|| \cdot ||n(x)||}. \tag{6.6}$$

For the RND $r(x)$ we have

$$q(r(x)) = -\frac{(\nabla f(x), r(x))}{||\nabla f(x)|| \cdot ||r(x)||}. \tag{6.7}$$

The following theorem establishes the lower bounds for $q(r(x))$ and $q(n(x))$. It shows that the regularization (6.1) improves the condition number of the Hessian $\nabla^2 f$ for all $x \in \mathbb{R}^n, x \notin X^*$ (see [10]).

Theorem 6.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuous differentiable convex function and the bounds (6.3) hold, then:*

1.
$$\begin{aligned} 1 \geq q(r(x)) &\geq (m(x) + ||\nabla f(x)||)(M(x) + ||\nabla f(x)||)^{-1} \\ &= \textit{cond } \ H(x) > 0 \qquad \textit{for any } x \notin X^*. \end{aligned}$$

2.
$$\begin{aligned} 1 \geq q(n(x)) \geq m(x)(M(x))^{-1} &= \textit{cond } \ \nabla^2 f(x) \\ &\textit{for any } x \in \mathbb{R}^n. \end{aligned}$$

3.
$$cond\ \ H(x) - cond\ \ \nabla^2 f(x) =$$
$$||\nabla f(x)||(1 - cond\ \ \nabla^2 f(x))(M(x) + ||\nabla f(x)||)^{-1} > 0 \tag{6.8}$$
for any $x \notin X^*$, $cond\ \ \nabla^2 f(x) < 1$.

Proof.

1. From (6.5), we obtain
$$||\nabla f(x)|| \leq ||H(x)|| \cdot ||r(x)||. \tag{6.9}$$
Using the right inequality (6.3) and 3°, we have
$$||H(x)|| \leq M(x) + ||\nabla f(x)||, \tag{6.10}$$
From (6.9) and (6.10) we obtain
$$||\nabla f(x)|| \leq (M(x) + ||\nabla f(x)||)||r(x)||.$$
From (6.5) the left inequality (6.3) and 3° follows
$$-(\nabla f(x), r(x)) = (H(x)r(x), r(x)) \geq (m(x) + \|\nabla f(x)\|)\|r(x)\|^2.$$
Therefore from (6.7) follows
$$q(r(x)) \geq (m(x) + ||\nabla f(x)||)(M(x) + ||\nabla f(x)||)^{-1} = \text{cond } H(x).$$
2. Now let us consider the Newton's direction $n(x)$. From (6.4), we have
$$\nabla f(x) = -\nabla^2 f(x) n(x), \tag{6.11}$$
therefore,
$$-(\nabla f(x), n(x)) = (\nabla^2 f(x) n(x), n(x)).$$
From (6.11) left inequality of (6.3), we obtain
$$q(n(x)) = -\frac{(\nabla f(x), n(x))}{||\nabla f(x)|| \cdot ||n(x)||} \geq m(x)||n(x)|| \cdot ||\nabla f(x)||^{-1}. \tag{6.12}$$
From (6.11) and the right inequality in (6.3) follows
$$||\nabla f(x)|| \leq ||\nabla^2 f(x)|| \cdot ||n(x)|| \leq M(x)||n(x)||. \tag{6.13}$$
Combining (6.12) and (6.13) we have
$$q(n(x)) \geq \frac{m(x)}{M(x)} = \text{cond } \nabla^2 f(x).$$
3. Using the formulas for the condition numbers of $\nabla^2 f(x)$ and $H(x)$ we obtain (6.3) □

Corollary 6.2. *The regularized Newton's direction $r(x)$ is a decent direction for any convex $f : \mathbb{R}^n \to \mathbb{R}$, whereas the classical Newton's direction $n(x)$ exists and it is a decent direction only if f is a strongly convex at $x \in \mathbb{R}^n$.*

Under condition (3.1) and (3.2) the RNM retains the local quadratic convergence rate, which is typical for the Classical Newton's method.

On the other hand, the regularization (6.1) allows to establish global convergence and estimate complexity of the RNM, when the original function is only strongly convex at x^*.

7. Local Quadratic Convergence Rate of the RNM

In this section we consider the RNM and determine the neighborhood of the minimizer, where the RNM converges with quadratic rate.

Along with assumptions (3.1) and (3.2) for the Hessian $\nabla^2 f$ we will use the Lipschitz condition for the gradient ∇f

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|, \tag{7.1}$$

which is equivalent to (4.5).

The RNM generates a sequence $\{x_s\}_{s=0}^{\infty}$:

$$x_{s+1} = x_s - \left[\nabla^2 f(x_s) + \|\nabla f(x_s)\| I\right]^{-1} \nabla f(x_s). \tag{7.2}$$

The following Theorem characterizes the regularized Newton's area.

Theorem 7.1. *If (3.1), (3.2) and* (7.1) *hold, then for* $\delta = \frac{2}{3}\frac{m}{M+2L}$ *and any* $x_0 \in B(x^*, \delta)$ *as a starting point, the sequence* $\{x_s\}_{s=0}^{\infty}$ *generated by RNM (7.2) belongs to* $B(x^*, \delta)$ *and the following bound holds:*

$$\Delta_{s+1} = \|x_{s+1} - x^*\| \le \frac{M+2L}{2} \cdot \frac{1}{m - (M+2L)\Delta_s}\Delta_s^2,\ s \ge 1. \tag{7.3}$$

Proof. From (7.2) follows

$$x_{s+1} - x^* = x_s - x^* - \left[\nabla^2 f(x_s) + \|\nabla f(x_s)\| I\right]^{-1} (\nabla f(x_s) - \nabla f(x^*)).$$

Using

$$\nabla f(x_s) - \nabla f(x^*) = \int_0^1 \nabla^2 f(x^* + \tau(x_s - x^*))(x_s - x^*)d\tau,$$

we obtain

$$x_{s+1} - x^* = \left[\nabla^2 f(x_s) + \|\nabla f(x_s)\| I\right]^{-1} H_s(x_s - x^*), \tag{7.4}$$

where

$$H_s = \int_0^1 (\nabla^2 f(x_s) + \|\nabla f(x_s)\| I - \nabla^2 f(x^* + \tau(x_s - x^*)))d\tau.$$

From (3.2) and (7.1) follows

$$\|H_s\| = \| \int_0^1 \left(\nabla^2 f(x_s) + \|\nabla f(x_s)\| I - \nabla^2 f(x^* + \tau(x_s - x^*))\right) d\tau \|$$

$$\leq \| \int_0^1 (\nabla^2 f(x_s) - \nabla^2 f(x^* + \tau(x_s - x^*))) d\tau \| + \int_0^1 \|\nabla f(x_s)\| d\tau$$

$$\leq \int_0^1 \|\nabla^2 f(x_s) - \nabla^2 f(x^* + \tau(x_s - x^*))\| d\tau + \int_0^1 \|\nabla f(x_s) - \nabla f(x^*)\| d\tau$$

$$\leq \int_0^1 M\|x_s - x^* - \tau(x_s - x^*)\| d\tau + \int_0^1 L\|x_s - x^*\| d\tau$$

$$= \int_0^1 (M(1-\tau) + L)\|x_s - x^*\| d\tau = \frac{M + 2L}{2}\|x_s - x^*\|. \tag{7.5}$$

From (7.4) and (7.5) we have

$$\Delta_{s+1} = \|x_{s+1} - x^*\| \leq \| \left(\nabla^2 f(x_s) + \|\nabla f(x_s)\| I\right)^{-1} \| \cdot \|H_s\| \cdot \|x_s - x^*\|$$

$$\leq \frac{M + 2L}{2} \|(\nabla^2 f(x_s) + \|\nabla f(x_s)\| I)^{-1}\| \Delta_s^2. \tag{7.6}$$

From (3.2) follows

$$\|\nabla^2 f(x_s) - \nabla^2 f(x^*)\| \leq M\|x_s - x^*\| = M\Delta_s, \tag{7.7}$$

therefore we have

$$\nabla^2 f(x^*) + M\Delta_s I \succeq \nabla^2 f(x_s) \succeq \nabla^2 f(x^*) - M\Delta_s I. \tag{7.8}$$

From (3.1) and (7.8) we obtain

$$\nabla^2 f(x_s) + \|\nabla f(x_s)\| I \succeq (m + \|\nabla f(x_s)\| - M\Delta_s) I.$$

Therefore for $\Delta_s < \frac{m + \|\Delta f(x_s)\|}{M}$ the matrix $\nabla^2 f(x_s) + \|\nabla f(x_s)\| I$ is positive definite, therefore its inverse exists and we have

$$\|(\nabla^2 f(x_s) + \|\nabla f(x_s)\| I)^{-1}\| \leq \frac{1}{m + \|\nabla f(x_s)\| - M\Delta_s} \leq$$

$$\frac{1}{m - M\Delta_s}. \tag{7.9}$$

For $\Delta_s \leq \frac{2}{3}\frac{m}{M+2L}$ from (7.6) and (7.9) follows

$$\Delta_{s+1} \leq \frac{M + 2L}{2} \frac{1}{m - (M + 2L)\Delta_s} \Delta_s^2. \tag{7.10}$$

Therefore from (7.10) for $0 < \Delta_s \leq \frac{2}{3}\frac{m}{M+2L} < \frac{m + \|\nabla f(x_s)\|}{M}$ we obtain

$$\Delta_{s+1} \leq \frac{3(M + 2L)}{2} \frac{1}{m} \Delta_s^2 \leq \Delta_s.$$

Hence, for $\delta = \frac{2}{3}\frac{m}{M+2L}$ and any $x_0 \in B(x^*, \delta)$ as a starting point the sequence $\{x_s\}_{s=0}^{\infty}$ generated by (7.2) belongs to $B(x^*, \delta)$ and the bound (7.3) holds.

The proof of Theorem 7.1 is completed. □

Corollary 7.2. *Under conditions of Theorem 7.1 for* $\delta = \frac{2}{3}\frac{m}{M+2L}$ *and any* $x \in B(x^*, \delta)$ *the Hessian* $\nabla^2 f(x)$ *is positive definite and*

$$\nabla^2 f(x) \succeq m_0 I, \tag{7.11}$$

where $m_0 = m(\frac{1}{3}M + 2L)(M + 2L)^{-1}$

In fact, from (4.4) follows

$$\nabla^2 f(x^*) + M\Delta x I \succeq \nabla^2 f(x) \succeq \nabla^2 f(x^*) - M\Delta x I,$$

so for any $x \in B(x^*, \delta)$ we have

$$\nabla^2 f(x) \succeq \left(m - \frac{2}{3}\frac{Mm}{M + 2L} \right) I = m(\frac{1}{3}M + 2L)(M + 2L)^{-1} I = m_0 I.$$

From the letter inequality follows

$$\|\nabla f(x)\| \|x - x^*\| \geq (\nabla f(x) - \nabla f(x^*), x - x^*) \geq m_0 \|x - x^*\|^2,$$

that is for any $x \in B(x^*, \delta)$ we have

$$\|\nabla f(x)\| \geq m_0 \|x - x^*\|. \tag{7.12}$$

It follows from Theorem 7.1 that $B(x^*, \delta)$ with $\delta = \frac{2}{3}\frac{m}{M+2L}$ is the Newton's area for the RNM.

So it takes $O(\ln \ln \varepsilon^{-1})$ regularized Newton's steps to find an ε-approximation for x^* from any $x \in B(x^*, \delta)$ as a starting point.

To make the RNM globally convergent we have to replace the RNM by DRNM and adjust the step length. It can be done by backtracking line search, using Armijo condition (5.13) with Newton's direction $n(x)$ replaced by regularized Newton's direction $r(x)$. In the following section we introduce another version of the DRNM and estimate the number of RNM steps required for finding $x \in \mathbb{B}(x^*, \delta)$ from any given starting point $x_0 \in \mathbb{R}^n$.

8. Damped Regularized Newton's Method

Let us consider the regularized Newton's decrement

$$\lambda_r(x) = (H^{-1}(x)\nabla f(x), \nabla f(x))^{\frac{1}{2}} = [-(\nabla f(x), r(x))]^{\frac{1}{2}}. \tag{8.1}$$

We assume that $\varepsilon > 0$ is small enough, in particular,

$$0 < \varepsilon^{0.5} < m_0 (L + \|\nabla f(x)\|)^{-0.5}, \tag{8.2}$$

for $\forall x \in \mathcal{L}_0$.

From (4.5) follows

$$(\nabla^2 f(x) + \|\nabla f(x)\| I) \preceq (L + \|\nabla f(x)\|) I. \tag{8.3}$$

On the other hand, for any $x \in B(x^*, \delta)$ from the Corollary 7.2 we have

$$\nabla^2 f(x) + \|\nabla f(x)\| I \succeq (m_0 + \|\nabla f(x)\|) I.$$

Therefore the inverse $(\nabla^2 f(x) + \|\nabla f(x)\| I)^{-1}$ exists and from (8.3) we obtain

$$H^{-1}(x) = (\nabla^2 f(x) + \|\nabla f(x)\| I)^{-1} \succeq (L + \|\nabla f(x)\|)^{-1} I.$$

Therefore from (8.1) for any $x \in B(x^*, \delta)$ we have

$$\lambda_{(r)}(x) = (H^{-1}(x)\nabla f(x), \nabla f(x))^{0.5} \geq (L + \|\nabla f(x)\|)^{-0.5}\|\nabla f(x)\|,$$

which together with (7.12) leads to

$$\lambda_{(r)}(x) \geq m_0(L + \|\nabla f(x)\|)^{-0.5}\|x - x^*\|.$$

Then from $\lambda_{(r)}(x) \leq \varepsilon^{1.5}$ and (8.2) follows

$$m_0(L + \|\nabla f(x)\|)^{-0.5}\varepsilon \geq \varepsilon^{1.5} \geq \lambda_{(r)}(x) \geq m_0(L + \|\nabla f(x)\|)^{-0.5}\|x - x^*\|$$

or

$$\|x - x^*\| \leq \varepsilon, \forall x \in B(x^*, \delta).$$

Therefore $\lambda_{(r)}(x) \leq \varepsilon^{1.5}$ can be used as a stopping criteria.

We are ready to describe the DRNM.

Let $x_0 \in \mathbb{R}^n$ be a starting point and $0 < \varepsilon < \delta$ be the required accuracy, set $x := x_0$.

1. Compute the regularized Newton's direction $r(x)$ by solving the system (6.5);
2. if the following inequality

$$f(x + tr(x)) \leq f(x) + 0.5(\nabla f(x), r(x)) \tag{8.4}$$

holds, then set $t(x) := 1$, otherwise set $t(x) := (2L)^{-1}\|\nabla f(x)\|$;

3. $x := x + t(x)r(x)$;
4. if $\lambda_r(x) \leq \varepsilon^{1.5}$, then $x^* := x$, otherwise go to 1.

The global convergence and the complexity of the DRNM we consider in the following section.

9. Complexity of the DRNM

We assume that conditions (3.1) and (3.2) are satisfied. Due to (3.1) the solution x^* is unique. Hence, from convexity f follows that for any given starting point $x_0 \in \mathbb{R}^n$ the sublevel set $\mathcal{L}_0$ is bounded, therefore there is $L > 0$ such that (4.5) holds on $\mathcal{L}_0$.

Let $\mathbb{B}(x^*, r) = \{x \in \mathbb{R}^n : \|x - x^*\| \leq r\}$ be the ball with center x^* and radius $r > 0$ and $r_0 = \min\{r : \mathcal{L}_0 \subset \mathbb{B}(x^*, r)\}$.

Theorem 9.1. *If (3.1) and (3.2) are satisfied and $\delta = \frac{2}{3}\frac{m}{M+2L}$, then from any given starting point $x_0 \in \mathcal{L}_0$ it takes*

$$N_0 = 13.5\left(\frac{L^2(M + 2L)^3}{(m_0 m)^3}(1 + r_0)(f(x_0) - f(x^*))\right) \tag{9.1}$$

DRN steps to get $x \in \mathbb{B}(x^*, \delta)$.

Proof. For the regularized Newton's directional derivative we have

$$\frac{df(x + tr(x))}{dt}|_{t=0} = (\nabla f(x), r(x)) =$$
$$-\left((\nabla^2 f(x) + \|\nabla f(x)\|I)r(x), r(x)\right) \leq$$

$$-(m(x)+\|\nabla f(x)\|)\|r(x)\|^2, \tag{9.2}$$

where $m(x)\geq 0$ and $\|\nabla f(x)\|>0$ for any $x\neq x^*$. It means that RND is a decent direction at any $x\in\mathcal{L}_0$ and $x\neq x^*$.

It follows from (9.2) that $\varphi(t)=f(x+tr(x))$ is monotone decreasing for small $t>0$.

From the convexity of f follows that $\varphi'(t)=(\nabla f(x+tr(x)),r(x))$ is not decreasing in $t>0$, hence at some $t=t(x)$ we have

$$(\nabla f(x+t(x)r(x)),r(x))\geq -\frac{1}{2}(m(x)+\|\nabla f(x)\|)\|r(x)\|^2, \tag{9.3}$$

otherwise $\inf f(x)=-\infty$, which is impossible due to the boundedness of $\mathcal{L}_0$.

From (9.2) and (9.3) we have

$$(\nabla f(x+t(x)r(x))-\nabla f(x),r(x))\geq \frac{m(x)+\|\nabla f(x)\|}{2}\|r(x)\|^2.$$

Therefore there exist $0<\theta(x)<1$ such that

$$t(x)(\nabla^2 f(x+\theta(x)t(x)r(x)),r(x))=t(x)(\nabla^2 f(\cdot)r(x),r(x))$$
$$\geq \frac{m(x)+\|\nabla f(x)\|}{2}\|r(x)\|^2$$

or

$$t(x)\|\nabla^2 f(\cdot)\|\|r(x)\|^2\geq \frac{m(x)+\|\nabla f(x)\|}{2}\|r(x)\|^2.$$

Keeping in mind that $\|\nabla^2 f(\cdot)\|\leq L$ we obtain

$$t(x)\geq \frac{m(x)+\|\nabla f(x)\|}{2L}\geq \frac{\|\nabla f(x)\|}{2L}. \tag{9.4}$$

It means that for $t\leq \frac{\|\nabla f(x)\|}{2L}$ the inequality

$$\frac{df(x+tr(x))}{dt}\leq -\frac{1}{2}(\nabla f(x),r(x))$$

holds, hence

$$\Delta f(x)=f(x)-f(x+t(x)r(x))\geq$$
$$\frac{1}{2}t(x)(-\nabla f(x),r(x))=\frac{1}{2}t(x)\lambda_r^2(x). \tag{9.5}$$

Therefore finding the lower bound for the decrease of f at any $x\in\mathcal{L}_0$ such that $x\notin B(x^*,\delta)$ we have to find the corresponding bound for the regularized Newton's decrement.

Now let us consider $x\in B(x^*,\delta)$ then from (7.11) follows

$$(\nabla f(x)-\nabla f(x^*),x-x^*)\geq m_0\|x-x^*\|^2. \tag{9.6}$$

for any $x\in B(x^*,\delta)$.

Let $\hat{x}\notin B(x^*,\delta)$, we consider a segment $[x^*,\hat{x}]$. There is $0<\widetilde{t}<1$ such that $\widetilde{x}=(1-\widetilde{t})x^*+\widetilde{t}\hat{x}\in\partial B(x^*,\delta)$.

From the convexity f follows

$$(\nabla f(x^*+t(\hat{x}-x^*)),\hat{x}-x^*)|_{t=0} \le (\nabla f(x^*+t(\hat{x}-x^*)),\hat{x}-x^*)|_{t=\widetilde{t}} \le (\nabla f(x^*+t(\hat{x}-x^*),\hat{x}-x^*)|_{t=1},$$

or

$$0=(\nabla f(x^*),\hat{x}-x^*) \le (\nabla f(\widetilde{x}),\hat{x}-x^*) \le (\nabla f(\hat{x}),\hat{x}-x^*).$$

The right inequality can be rewritten as follows:

$$(\nabla f(\widetilde{x}),\hat{x}-x^*)=\frac{\|\hat{x}-x^*\|}{\delta}(\nabla f(\widetilde{x})-\nabla f(x^*),\widetilde{x}-x^*) \le (\nabla f(\hat{x}),\hat{x}-x^*).$$

In view of (9.6) we obtain

$$\|\nabla f(\hat{x})\|\|\hat{x}-x^*\| \ge \frac{\|\hat{x}-x^*\|}{\delta}(\nabla f(\widetilde{x})-f(x^*),\widetilde{x}-x^*) \ge \frac{\|\hat{x}-x^*\|}{\delta}m_0\|\widetilde{x}-x^*\|^2.$$

Keeping in mind that $\widetilde{x}\in\partial B(x^*,\delta)$ we get

$$\|\nabla f(\hat{x})\| \ge m_0\|\widetilde{x}-x^*\| = \frac{2}{3}m_0 m\frac{1}{M+2L}. \tag{9.7}$$

On the other hand from (7.1) and $\hat{x}\in\mathcal{L}_0$ follows

$$\|\nabla f(\hat{x})\| = \|\nabla f(\hat{x})-\nabla f(x^*)\| \le L\|\hat{x}-x^*\| \le Lr_0. \tag{9.8}$$

From (4.5) follows

$$\nabla^2 f(x) \preceq LI. \tag{9.9}$$

For any $\hat{x}\notin S(x^*,\delta)$ we have $\|\nabla f(\hat{x})\|>0$, therefore $H(\hat{x})=\nabla^2 f(\hat{x})+\|\nabla f(\hat{x})\|I$ is positive definite and system (6.5) has a unique solution

$$r(\hat{x})=-H^{-1}(\hat{x})\nabla f(\hat{x}).$$

Moreover from (9.9) follows

$$(\nabla^2 f(\hat{x})+\|\nabla f(\hat{x})\|I) \preceq (L+\|\nabla f(\hat{x})\|)I.$$

Therefore

$$H^{-1}(\hat{x}) \succeq (L+\|\nabla f(\hat{x})\|I)^{-1})I. \tag{9.10}$$

For the regularized Newton's decrement we obtain

$$\lambda_{(r)}(\hat{x})=(H^{-1}(x))\nabla f(\hat{x}),\nabla f(\hat{x}))^{0.5} \ge (L+\|\nabla f(\hat{x}\|)^{-0.5}\|\nabla f(\hat{x})\|. \tag{9.11}$$

Keeping in mind

$$\|\nabla f(\hat{x}\| = \|\nabla f(\hat{x})-\nabla f(x^*)\| \le L\|\hat{x}-x^*\|$$

from (9.4), (9.8) and (9.11) and definition of r_0 we obtain

$$\Delta f(\hat{x}) \ge \frac{1}{2}t(\hat{x})\lambda_r^2(\hat{x}) \ge \frac{\|\nabla f(\hat{x})\|^3}{4L}(L+\|\nabla f(\hat{x})\|)^{-1} \ge \frac{\|\nabla f(\hat{x})\|^3}{4L^2(1+r_0)}.$$

Using (9.7) we get

$$\Delta f(\hat{x}) \ge \left(\frac{2}{3}m_0 m\frac{1}{M+2L}\right)^3\frac{1}{4L^2(1+r_0)}$$

$$= \frac{2}{27}\frac{(m_0 m)^3}{(M+2L)^3L^2}\frac{1}{(1+r_0)}.$$

Therefore it takes

$$N_0 = (f(\hat{x}) - f(x^*))\Delta f^{-1}(\hat{x}) = 13.5\frac{(M+2L)^3L^2}{(m_0 m)^3}(1+r_0)(f(\hat{x}) - f(x^*))$$

steps to obtain $x \in \mathbb{B}(x^*, \delta)$ from any $x \in \mathcal{L}_0$.

The proof of Theorem 9.1 is completed. □

From (7.3) follows that it takes $O(\ln\ln\varepsilon^{-1})$ DRN steps to find an ε-approximation for x^* from any $x \in \mathbb{B}(x^*, \delta)$.

Therefore the total number of DRN steps required for finding an ε-approximation for x^* from a given starting point $x_0 \in \mathbb{R}^n$ is

$$N = N_0 + o(\ln\ln\varepsilon^{-1}).$$

10. Concluding Remarks

The bounds (5.1) and (9.1) depends on the size of Newton's and regularized Newton's areas, which, in turn, are defined by convexity constant $m > 0$ and smoothness constants $M > 0$ and $L > 0$. The convexity and smoothness constants dependent on the given system of coordinate.

Let consider an affine transformation of the original system given by $x = Ay$, where $A \in \mathbb{R}^{n\times n}$ is a nondegenerate matrix. We obtain $\varphi(y) = f(Ay)$.

Let $\{x_s\}_{s=0}^{\infty}$ be the sequence generated by Newton's method

$$x_{s+1} = x_s - (\nabla^2 f(x_s))^{-1}\nabla f(x_s).$$

For the correspondent sequence in the transformed space we obtain

$$y_{s+1} = y_s - (\nabla^2\varphi(y_s))^{-1}\nabla\varphi(y_s).$$

Let $y_s = A^{-1}x_s$ for some $s \geq 0$, then

$$y_{s+1} = y_s - (\nabla^2\varphi(y_s))^{-1}\nabla\varphi(y_s) = y_s - [A^T\nabla^2 f(Ay_s)A]^{-1}A^T\nabla f(Ay_s) =$$
$$A^{-1}x_s - A^{-1}(\nabla^2 f(x_s))^{-1}\nabla f(x_s) = A^{-1}x_{s+1}.$$

It means that Newton's method is affine invariant with respect to the transformation $x = Ay$. Therefore the areas of quadratic convergence depends only on the local topology of f(see [7]).

To get the Newton's sequence in the transformed space one needs to apply A^{-1} to the elements of the Newton's sequence in the original space.

Let N is such that $x_N : \|x_N - x^*\| \leq \varepsilon$, then

$$\|y_N - y^*\| \leq \|A^{-1}\|\|x_N - x^*\|.$$

From (3.3) follows

$$\|x_{N+1} - x^*\| \leq \frac{M}{2(m - M\|x_s - x^*\|)}\|x_N - x^*\|^2.$$

Therefore

$$\|y_{N+1} - y^*\| \leq \|A^{-1}\|\|x_{N+1} - x^*\| \leq \frac{1}{2}\|A^{-1}\|\frac{M}{(m - M\varepsilon)}\varepsilon^2.$$

Hence, for small enough

$$\varepsilon \leq 0.5\frac{m}{M}\min\{1; (\|A^{-1}\|)^{-1}\}$$

we have

$$\|y_{N+1} - y^*\| \leq \varepsilon.$$

We would like to emphasize that the bound (9.1) is global, while the conditions (3.1) and (3.2) under which the bound holds are local, at the neighborhood of x^*.

References

[1] V. Arnold, *Small denominators and problem of stability in classical and celestial mechanics*, Uspehi Matematicheskih Nauk 18(6), (1963) 91-192.
[2] S. Boyd and Vandenberghe, *Convex Optimization*, Cambridge University Press. (2004).
[3] A. Ioffe, *On the local surjection property.* Nonlinear Analysis, VII, n 5 (1987),565–592.
[4] L. Kantorovich and Akilow, *Functional Analysis*, Nauka, Moscow (1977).
[5] Ljusternik, *On conditional extrema of functionals*, Mathematicheskij Sbornik 41(3), (1934), 390–401.
[6] Yu. Nesterov and A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia (1994).
[7] Yu. Nesterov, *Introductory Lectures on Convex Programming*, Kluwer, Boston (2004).
[8] B. Polyak, *Introduction to Optimization*, Optimization Software, Inc. NY, (1987).
[9] B. Polyak *Newton's method and its use in Optimization*, European Journal of OR, 181 (2007) 1086-1096 .
[10] R. Polyak, *Regularized Newton method for unconstrained Convex Optimization*, Mathematical programming Ser. B 120, (2009), 125-145.

(R. Polyak) Department of Mathematics, The Technion - Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: rpolyak@techunix.technion.ac.il and rpolyak@gmu.edu