Convergence of Newton's method on self-concordant convex functions

Srikanth Avasarala

1 Introduction

Newton's method has been recognized as an effective method for optimizing convex functions on $\mathbb{R}^n$ owing to its fast quadratic local convergence. In this expository note, we dilineate on the classical convergence analysis of Newton's method on Convex functions, and its drawbacks in providing 'scale-independent' convergence guarantees. To circumvent the above problem, Nesterov and Nemirovsky introduced the notion of Self-concordance, a constraint set on convex functions to enable scale-free, affine-invariant convergence analysis. In this expository note, we highlight on the classical convergence guarantees of Newton's method, and then move on to the convergence analysis of Newton's method for self-concordant convex functions. To verify the nature of convergence bounds, we perform numerical simulations on a toy convex self-concordant problem.

2 Newton's method of optimization

2.1 Theory

Newton's method falls under the class of line-search methods for optimizing convex functions. Line search methods follow a standard protocol of computing a search direction $\Delta x_{nt} (\in \mathbb{R}^n)$ and then deciding on how far to move along that direction, given by $t (\in \mathbb{R})$. The iteration at step k is given by

$$x^+ = x + t\Delta x_{nt} \tag{1}$$

In the Newton's method, $\Delta x_{nt} = \nabla^2 f(x)^{-1} \nabla f(x)$ where Δx_{nt} is called the Newton's step. From the positive definiteness of $\nabla^2 f(x)$ we have

$$-\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0 \tag{2}$$

until $\nabla f(x) = 0$ (which is where the function is minimized). The quantity $(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2} := \lambda(x)$ is called the Newton Decrement of x. We now show that Newton's decrement is an important quantity as it is equal to $f(x) - \inf_y \hat{f}(y)$ where $\hat{f}(y)$ is a quadratic approximation of f at y.

$$f(x) - \inf_y \hat{f}(y) = f(x) - min_y(f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x)$$

$$= f(x) - (f(x) - \frac{1}{2} \nabla f(x)^T (\nabla^2 f(x))^{-1} \nabla f(x))$$

$$= \frac{1}{2}\lambda(x)^2 \tag{3}$$

The above equation shows that $\lambda(x)^2$ can be used as an approximate bound on the suboptimality gap $f(x) - p^*$.

2.1.1 Backtracking line search

Newton's method need not converge under fixed step length. Backtracking line search is an iterative method used to choose the step size t at a given iteration. Setting parameters $0 < \alpha \leq 1/2$, $0 < \beta < 1$. At each iteration, we start with $t = 1$ and **while**

$$f(x + tv) > f(x) + \alpha t \nabla f(x)^T v$$

do $t = \beta t$. We perform Newton's update with the output value of t once we come out of the loop. Here $v = -(\nabla^2 f(x))^{-1} \nabla f(x)$, so $\nabla f(x) v = -\lambda^2(x)$.

2.2 Convergence analysis of Newton's method

To perform the convergence analysis, we assume that f is convex, twice-differentiable, having $dom f = \mathbb{R}^n$. Additionally we assume $\nabla^2 f$ is L-Lipschitz i.e $||\nabla^2 f(x) - \nabla^2 f(y)||_2 \leq L||x - y||_2$, f is m-strongly convex. As a consequence of strong convexity of f, we have $mI \preceq \nabla^2 f(x) \preceq MI$ where $M > 0$. Given an error tolerance of ε, we seek to find a vector $\tilde{x}$ such that $f(\tilde{x}) - p^* \leq \varepsilon$. We now state the theorem for the convergence analysis of Newton's method without providing the proof (as it is very similar to the one that we will perform for self-concordant functions later on).

THEOREM 1. Newton's method with backtracking line search with line-search parameters α, β satisfies the following two-stage convergence bounds

$$f(x^k) - p^* \leq \begin{cases} (f(x_0) - p^*) - \gamma k, & \text{if } k \leq k_0. \\ \frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{k-k_0+1}}, & k > k_0\ . \end{cases}$$

Here $\gamma = \alpha\beta\eta^2 m/M^2$, $\eta = min 1, 3(1 - 2\alpha)m^2/L$ and k_0 is the number of steps of iterations until $||\nabla f(x^{k_0+1})||_2 < \eta$.

Elaborating further on the above theorem, there exists numbers $0 < \eta \leq m^2/L$ and $\gamma > 0$ such that the following hold.

- If $||\nabla f(x^k)||_2 \geq \eta$, then
$$f(x^{k+1}) - f(x^k) \leq -\gamma \tag{4}$$
- If $||\nabla f(x^k)||_2 < \eta$, then the backtracking line search selects $t^k = 1$ and
$$\frac{L}{2m^2}||\nabla f(x^{k+1})||_2 \leq \left(\frac{L}{2m^2}||\nabla f(x^k)||_2\right)^2 \tag{5}$$

Let us interpret the implications of the second condition. From eq. 5, it is easy to see that $||\nabla f(x^{k+1})||_2 < \eta$ is satisfied for the iteration $k+1$ given that eq. 5 is satisfied for iteration k. Therefore the second condition holds for all future iterates i.e for all $l \geq k$ and we have

$$\frac{L}{2m^2}||\nabla f(x^{l+1})||_2 \leq \left(\frac{L}{2m^2}||\nabla f(x^l)||_2\right)^2 \tag{6}$$

By induction, we have for $l \geq k$,

$$\frac{L}{2m^2}||\nabla f(x^l)||_2 \leq \left(\frac{L}{2m^2}||\nabla f(x^l)||_2\right)^{2^{l-k}} \leq \left(\frac{1}{2}\right)^{2^{l-k}} \tag{7}$$

From Polyak's lemma for m-strongly convex functions (discussed in the class), we have

$$f(x^l) - p^* \leq \frac{1}{2m}||\nabla f(x^l)||_2^2 \leq \frac{2m^3}{L^2}\Big(\frac{1}{2}\Big)^{2^{l-k+1}} \tag{8}$$

The above inequality shows that convergence is rapid once the second condition is satisfied. This phenomenon is called *quadratic convergence*. The iterations in Newton's method fall in two stages. The first one us refered to as *damped Newton phase* (due to the Newton step size t being less than 1), and the second stage where $||\nabla f(x)||_2 < \eta$ holds, is called the *quadratically convergent phase*.

We now estimate the total complexity in terms of the bounds on the number of iterations. Considering the first stage, i.e the damped Newton phase first, we see from eq. 4, the number of iterations is less than

$$\frac{f(x^0) - p^*}{\gamma}$$

To bound the number of iterations in the quadratically convergent phase, we use eq. 8 to show that under error tolerance ε doesn't exceed

$$log_2 log_2(\varepsilon_0/\varepsilon)$$

iterations in the quadratically convergent phase, where $\varepsilon_0 = 2m^3/L^2$. The overall number of iterations until $f(x) - p^* \leq \varepsilon$ is bounded above by

$$\frac{f(x^0) - p^*}{\gamma} + log_2 log_2(\varepsilon_0/\varepsilon)$$

where $\gamma = \alpha\beta\eta^2 m/M^2$, $\varepsilon_0 = 2m^3/L^2$, and $\eta = min\{1, 3(1-2\alpha)\}m^2/L$. The second term $log_2 log_2(\varepsilon_0/\varepsilon)$ in the above expression can be taken as a constant for all practical purposes, say five or six (Six iterations of a quadratically convergent phase gives an accuracy $\varepsilon \approx 5 \times 10^{-20}\varepsilon_0$).

2.3 Drawbacks of convergence analysis via Classical Newton's method

The traditional results on the Newton method state under reasonable smoothness and nondegeneracy assumptions lead to its local quadratic convergence. These results possess a generic conceptual drawback: the quantitive description of the region of quadratic convergence is given in terms of the condition number of the Hessian of f at the minimizer and the Lipschitz constant of this Hessian. These quantities, anyhow, are dependent on the choice of metric used: they are defined not by f itself, but also by the Euclidean structure in the space of variables. Another shortcoming of classical convergence analysis is a practical one. The complexity estimates obtained are a function of constants m, M, and L, which are not known in practice. Also the choice of the above constants change upon changing the coordinates of the space (Note the dependence of the constants on the Euclidean norm). We now seek a method of convergence analysis that is invariant to affine coordinate transforms. Therefore, we look for an alternative to the following conditions on f

- $mI \preceq \nabla^2 f(x) \preceq MI$ where $M > 0$ (Strong convexity).
- $||\nabla^2 f(x) - \nabla^2 f(y)||_2 \leq L||x - y||_2$ (L-Lipschitz property of Hessian)

The alternative constraints on f were imposed by Nesterov and Nemirovsky, called as self-concordance, that achieves the goal. In the next section, we introduce the notion of self-concordant functions, thier properties and the underlying calculus, followed by their convergence analysis under Newton's method.

3 Self-Concordance

3.1 Self-Concordant functions

A convex function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is *self-concordant* if

$$|f^{'''}(x)| \leq 2f^{''}(x)^{3/2} \tag{9}$$

Note that one can use an absolute constant κ other than 2 in the definition.

Examples

- Linear and Quadratic functions are self-concordant
- $f(x) = -ln(x), x > 0$ is self-concordant: $f^{''}(x) = \frac{1}{x^2}, f^{'''}(x) = -\frac{2}{x^3}, \frac{|f^{'''}(x)|}{f^{''}(x)^{3/2}} = 2, x > 0$
- $f(x) = e^x$ is NOT self-concordant as $\frac{|f^{'''}(x)|}{f^{''}(x)^{3/2}} = 2 \longrightarrow \infty$ as $x \longrightarrow -\infty$.

Scaling and affine invariance
Given that f is self-concordant and define $\tilde{f}(y) = ay + b\ (a \neq 0)$, then $\tilde{f}$ is concordant with the same κ.

Self-concordant function in $\mathbb{R}^n$
A convex function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is self-concordant if $g(t) = f(x + tv)$ is self-concordant for every $x \in \mathbb{R}^n, v \in \mathbb{R}^n$. Note that the constant κ of self-concordance is invariant to the choice of x and v.

Self-concordance preserving operations
The following are some important operations that preserve self-concordance.

- Scaling and sum: When f is self-concordant, $af\ (a > 1)$, $f_1 + f_2$ are both self-concordant
- Affine transformation: f is self-concordant,then $f(Ax + b)$ is self-concordant $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^n$

3.2 Implications of Self-concordance: Suboptimality bounds

Earlier, we used strong convexity and L-Lipschitz property of Hessians to derive the suboptimality bounds in terms of $||\nabla f(x)||$. We now use the Newton's decrement $\lambda(x)$ to attain similar bounds for strictly convex self-concordant functions.

$$(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2} := \lambda(x)$$

Note that $\lambda(x) > 0$ at every point x which is a non-minimizer of f due to the positive definiteness of $\nabla^2 f(x)$. It can be seen that for any $v \in \mathbb{R}^n$:

$$\lambda(x) = sup_{v \neq 0} \frac{-v^T \nabla f(x)}{(v^T \nabla^2 f(x) v)^{1/2}} \tag{10}$$

with the supremum attained at $v = -[\nabla^2 f(x)]^{-1} \nabla f(x) = \Delta x_{nt}$ (Easy to check by computing grad and Hessian of RHS wrt v).

3.2.1 Bounds on second derivative of f

Given a strictly convex self-concordant function $f : \mathbb{R} \longrightarrow \mathbb{R}$, we can write the self-concordant condition (eq. 9) as

$$\left|\frac{d}{dx}\left(f^{''}(x)^{-1/2}\right)\right| \leq 1 \tag{11}$$

for all $x \in \mathbb{R}$. Assume some $t > 0$. Integrating the above inequality both sides from 0 to t gives

$$-t \leq \int_0^t \frac{d}{dx}\left(f^{''}(x)^{-1/2}\right)dx \leq t \tag{12}$$

Using Leibnitz rule of integration, we get the following upper and lower bounds on $f^{''}(t)$

$$\frac{f^{''}(0)}{\left(1 + t\sqrt{f''(0)}\right)^2} \leq f^{''}(t) \leq \frac{f^{''}(0)}{\left(1 - t\sqrt{f''(0)}\right)^2} \tag{13}$$

where the lower bound is valid for $t \geq 0$ and the upper bound is valid for $0 \leq t < f''(0)^{-1/2}$.

3.2.2 Bounds on f

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a strictly convex self concordant function. Let v be a descent direction (satisfies $\nabla f(x)^T v < 0$). Consider $g(t) = f(x + tv)$, a self-concordant function by definition. Plugging g into the lower bound in eq. 13, we get a lower bound on $g^{'}(t)$

$$g^{'}(t) \geq g^{'}(0) + g^{''}(0)^{1/2} - \frac{g^{''}(0)^{1/2}}{1 + tg^{''}(0)^{1/2}} \tag{14}$$

Integrating the above inequality again gives a lower bound on $g(t)$:

$$g(t) \geq g(0) + tg^{'}(0) + tg^{''}(0)^{1/2} - log(1 + tg^{''}(0)^{1/2}) \tag{15}$$

The RHS in the above inequality is minimized at

$$t_0 = \frac{-g^{'}(0)}{g^{''}(0) + g^{''}(0)^{1/2}g^{'}(0)},$$

Taking infimum wrt $t \geq 0$ both sides of eq. 15, we have

$$inf_{t\geq 0}g(t) \geq g(0) - g^{'}(0)g^{''}(0)^{-1/2} + log(1 + g^{'}(0)g^{''}(0))^{-1/2} \tag{16}$$

Note that we substituted t_0 in RHS of eq. 15 to get the RHS of the above inequality. On the other hand, the inequality 10 can be expressed as

$$\lambda(x) \geq -g^{''}(0)^{-1/2}g^{'}(0)$$

Using the fact that $s + log(1 - s)$ is monotonically decreasing in s we get

$$inf_{t\geq 0}g(t) \geq g(0) + \lambda(x) + log(1 - \lambda(x))$$

The above inequality holds for any gradient direction v. Hence

$$p^* \geq f(x) + \lambda(x) + log(1 - \lambda(x)) \tag{17}$$

holds for $\lambda(x) < 1$. Using the properties of the function $-(\lambda(x) + log(1 - \lambda(x)))$, we have

$$-(\lambda(x) + log(1 - \lambda(x))) \approx \lambda^2/2$$

and the bound

$$-(\lambda(x) + log(1 - \lambda(x))) \leq \lambda^2$$

for $\lambda(x) < 0.68$. Thus we have the following bound on the sub-optimality of f

$$p^* \geq f(x) - \lambda(x)^2 : \ \lambda(x) < 0.68 \tag{18}$$

We earlier used $\lambda(x)^2/2$ as a stopping criterion for Newton's method (eq. 3). Now we have shown that for self-concordant functions, the stopping criterion

$$\lambda^2(x) \leq \varepsilon$$

where $\varepsilon < 0.68^2$, guarantees that on the exit $f(x) - p^* \leq \varepsilon$.

3.3 Convergence analysis of Newton's method for Self-concordant functions

Everything that we have built so far leads to this. We now analyze Newton's method with backtracking line search (with parameters α, β) to optimize a strictly convex self-concordant function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. We assume a starting point x^0.

THEOREM 2. Newton's method with backtracking line search (parameters being α, β) on a **strictly convex self-concordant function**, $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, requires atmost

$$C(\alpha, \beta)(f(x^0) - p^*) + loglog(1/\varepsilon)$$

iterations to reach $(f(x^k) - p^*) \leq \varepsilon$, where $C(\alpha, \beta) = \frac{20-8\alpha}{\alpha\beta(1-2\alpha)^2}$, $0 < \alpha < 1/2$, $0 < \beta < 1$

In the process of proving the theorem, we also show that there exists $\eta, \gamma > 0$, where $0 < \eta < 1/4$, that depend only on the parameters α, β, that lead to the following two phases of Convergence:

- If $\lambda(x^k) > \eta$, then
$$f(x^{k+1}) - f(x^k) \leq -\gamma \tag{19}$$
- If $\lambda(x^k) \leq \eta$, then backtracking line search selects $t = 1$ and,
$$2\lambda(x^{k+1}) \leq \left(2\lambda(x^{k+1})\right)^2 \tag{20}$$

Observe that the above two phases are very much similar to the one in classical Newton's convergence analysis that we discussed in subsection 2.2. The difference between both of them arises in the use of concordance assumption, instead of Hessian Lipschitzness and strong convexity. Also, the Newton decrement here $\lambda(x)$ will play the role of norm of gradient of f.

For the second case, we write down the resultant inequalities for $l \geq k$, where we have $\lambda(x^k) \leq \eta$ (like we did for 7, 8).

$$2\lambda(x^l) \leq (2\lambda(x^k))^{2^{l-k}} \leq (2\eta)^{2^{l-k}} \leq \left(\frac{1}{2}\right)^{2^{l-k}} \tag{21}$$

Hence for $l \geq k$,

$$f(x^l) - p^* \leq \lambda(x^l)^2 \leq \frac{1}{4}\left(\frac{1}{2}\right)^{2^{l-k+1}} \leq \left(\frac{1}{2}\right)^{2^{l-k+1}} \tag{22}$$

We recycle both the terms Damped Newton phase and Quadratically convergent phase, that we earlier used for classical analysis, to describe likewise the stages of convergence in case of self-concordant functions.

The total number of iterations that guarantee the accuracy $f(x) - p^* \leq \varepsilon$, starting from point x^0 does not exceed

$$\frac{(f(x^0) - p^*)}{\gamma} + loglog(1/\varepsilon)$$

iterations (Similar proof as seen in subsection 2.2).

We now proceed to prove the Theorem 2 by completing the convergence analysis in both phases of Newton's iterations.

3.3.1 Damped Newton phase

We now make use of the properties of self-concordant functions that we discussed in subsections 3.2.1, 3.2.2. Let $g(t) = f(x + t\Delta x_{nt})$. From integrating the upper bound of ineq.13 twice, we have

$$g(t) \leq g(0) + tg^{'}(0) - tg^{''}(0)^{1/2} - log(1 - tg^{''}(0)^{1/2})$$

$$= g(0) - t\lambda(x)^2 - t\lambda(x) - log(1 - t\lambda(x)) \tag{23}$$

The above inequality is valid for $0 \leq t < 1/\lambda(x)$. Now consider a point $\hat{t} = 1/(1 + \lambda(x))$. Note that this point satisfies the exit condition of backtracking line search. We now use the bound above to show that the step size $t \geq 1/(1 + \lambda(x))$ satisfies the exit condition of line search.

$$g(t) \leq g(0) - t\lambda(x)^2 - t\lambda(x) - log(1 - t\lambda(x))\,, \tag{24}$$

$$= g(0) - \lambda(x) + log(1 + \lambda(x))\,. \tag{25}$$

Using the following inequality

$$-x + log(1 + x) + \frac{x^2}{2(1 + x)} \leq 0 \tag{26}$$

where $x \geq 0$. Therefore, we get

$$g(t) \leq g(0) - \alpha\frac{\lambda(x)^2}{1 + \lambda(x)}\,, \tag{27}$$

$$= g(0) - \alpha\lambda(x)^2\hat{t}\,. \tag{28}$$

Since $t \geq 1/(1 + \lambda(x))$, we have

$$g(t) - g(0) \leq -\alpha\beta\frac{\lambda(x)^2}{1 + \lambda(x)} \tag{29}$$

Transforming g to f and using the condition $\lambda(x) > \eta$, we prove 19 with

$$\gamma = \alpha\beta\frac{\eta^2}{1 + \eta}$$

3.3.2 Quadratically convergent phase

Under the condition $\lambda(x) \leq \eta$, we can take

$$\eta = \frac{1-2\alpha}{4}$$

Since $0 < \alpha < 1/2$, the condition $0 < \eta < 1/4$ holds. In other words, if $\lambda(x^k) \leq (1-2\alpha)/2$ holds (as it does under the above conditions), then backtracking line search accepts the unit step $t = 1$ and (20) holds. It is important to note that $t = 1$ leads to a point in the domain of RHS of (23) if $\lambda(x) < 1$. Also, under the condition $\lambda(x) \leq (1-2\alpha)/2$, (23) implies

$$g(1) \leq g(0) - \lambda(x)^2 - \lambda(x) - log(1-\lambda(x)) \tag{30}$$

Using the fact that $-z - log(1-z) \leq \frac{1}{2}z^2 + z^3$ for $0 \leq z \leq 0.81$. We have for $z = \lambda(x)$ the following

$$g(1) \leq g(0) - \frac{1}{2}\lambda(x)^2 + \lambda(x)^3\,, \tag{31}$$

$$\leq g(0) - \alpha\lambda(x)^2 \tag{32}$$

which proves that unit step satisfies sufficient decrease condition to exit the backtracking loop. Now, given that $\lambda(x) < 1$, and $x^+ = x - \nabla^2 f(x)^{-1}\nabla f(x)$ (Newton's descent). We use the following property of self-concordant function f

$$(1-t\alpha)^2\nabla^2 f(x) \preceq \nabla^2 f(x+tv) \preceq \frac{1}{(1-t\alpha)^2}\nabla^2 f(x) \tag{33}$$

where $0 \leq t < \alpha$, $\alpha = (v^T\nabla^2 f(x)v)^{1/2}$; we show that

$$\lambda(x^+) \leq \frac{\lambda(x)^2}{(1-\lambda(x))^2} \tag{34}$$

Therefore, for $\lambda(x) \leq 1/4$

$$\lambda(x^+) \leq 2\lambda(x)^2 \tag{35}$$

thus proving (20) under $\lambda(x^k) \leq \eta\ (< 1/4)$.

3.3.3 Summing up for the final bound

The bound on total number of iterations required for ε error convergence becomes

$$\frac{(f(x^0)-p^*)}{\gamma} + loglog(1/\varepsilon) = \frac{20-8\alpha}{\alpha\beta(1-2\alpha)^2}(f(x^0)-p^*) + loglog(1/\varepsilon) \tag{36}$$

The coefficient of $(f(x^0)-p^*)$ in the RHS of the above equation is obtained by by plugging in $\gamma = \alpha\beta\frac{\eta^2}{1+\eta}$, $\eta = \frac{1-2\alpha}{4}$. This proves the Theorem 2.

We see that the final complexity bound depends only on the line search parameters α, β unlike the classical convergence which depends on smoothness, convexity parameters and the Euclidean nature of the space.

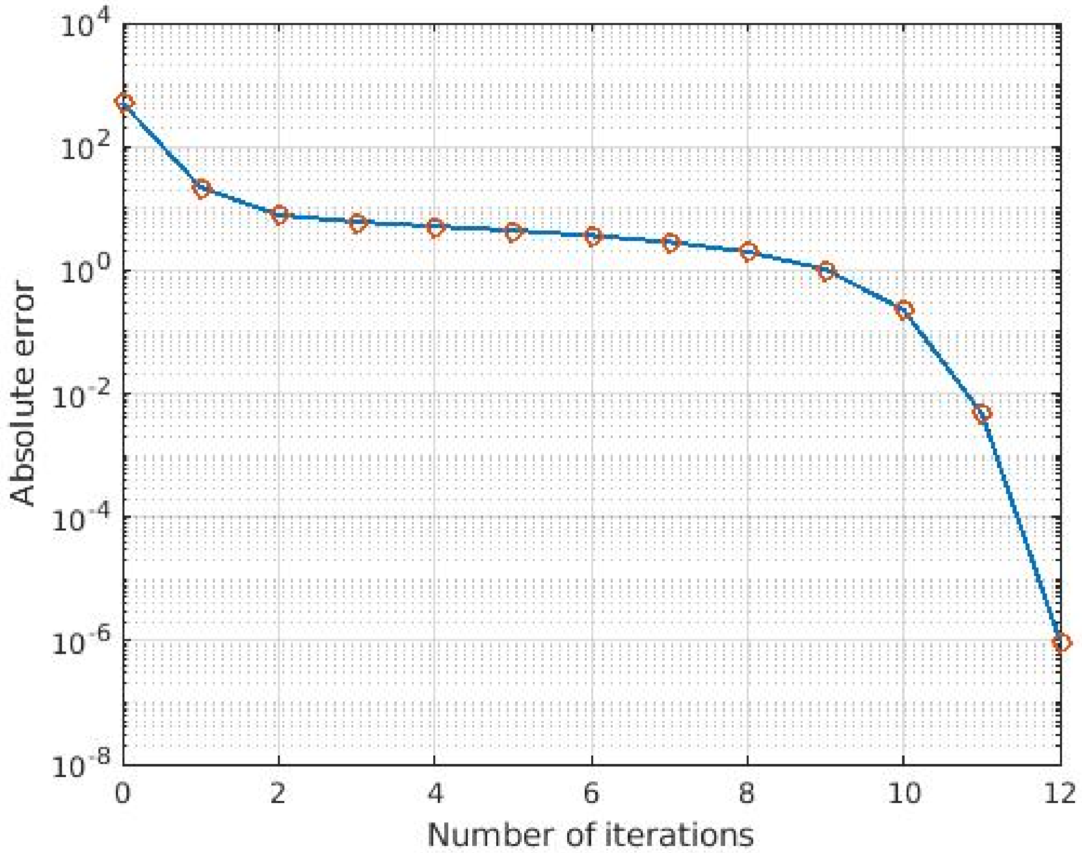

Figure 1: Absolute error vs number of iterations under Newton's scheme

4 Numerical verification of convergence on a self-concordant function

We choose the following self-concordant function, stated as an example of a self-concordant barrier function by Nesterov and Nemirovsky in their expository article, to optimize by using Newton's search.

$$f(x) = -\sum_{i=1}^{m} log(1 - {a_i}^T x) - \sum_{i=1}^{m} log(1 - x^2) \tag{37}$$

for $x \in dom f \subset \mathbb{R}^n$. We generate m samples of a_i, i = 1,2,...m, from a standard Normal distribution. In this problem setting, we choose $m = 10000, n = 1000$. $x_0 = 0 \in dom f$ is chosen as the starting point. The Newton backtracking parameters chosen are $\alpha = 0.01, \beta = 0.5$. The error tolerance ε is set to be 10^{-8}. We then plot the semi log plot of absolute error vs numver of iterations (until convergence). We notice that the plot depicts a nearly constant slope until the 9th iteration (damped phase) and then drops gradually after the 9th (Quadratic convergence phase). In total, 14 iterations (9 damped + 5 quadratic) sufficed to attain the given tolerance. We also compared the number of iterations to the convergence bound obtained in Theorem 2, but in the set of all computationally feasible parameters that we could take up for this problem, the convergence bound is too loose ($10^4 - 10^5$ order). Nevertheless the bounds are only dependent on backtracking parameters and error tolerance and are invariant to scaling or any other affine transformation performed to the function, which is the essence of self-concordance.

5 References

1. Nesterov, Y., Nemirovskii, A. (1994). Interior-point polynomial algorithms in convex programming. Society for industrial and applied mathematics.

2. Boyd, S., Boyd, S. P., Vandenberghe, L. (2004). Convex optimization. Cambridge university press.

3. Nocedal, J., Wright, S. J. (Eds.). (1999). Numerical optimization. New York, NY: Springer New York.

5. Nemirovski, A. (2004). Interior point polynomial time methods in convex programming. Lecture notes, 42(16), 3215-3224.