New Results on Superlinear Convergence of Classical Quasi-Newton Methods

Anton Rodomanov · Yurii Nesterov

Abstract We present a new theoretical analysis of local superlinear convergence of classical quasi-Newton methods from the convex Broyden class. As a result, we obtain a significant improvement in the currently known estimates of the convergence rates for these methods. In particular, we show that the corresponding rate of the Broyden–Fletcher–Goldfarb–Shanno method depends only on the product of the dimensionality of the problem and the *logarithm* of its condition number.

1 Introduction

We study local superlinear convergence of classical quasi-Newton methods for smooth unconstrained optimization. These algorithms can be seen as an approximation of the standard Newton method, in which the exact Hessian is replaced by some operator, which is updated in iterations by using the gradients of the objective function. The two most famous examples of quasi-Newton algorithms are the Davidon–Fletcher–Powell (DFP) [1, 2] and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) [3–7] methods, which together belong to the Broyden family [8] of quasi-Newton algorithms. For an introduction into the topic, see [9] and [10, Chapter 6]. See also [11] for the discussion of quasi-Newton algorithms in the context of nonsmooth optimization.

Anton Rodomanov
ICTEAM, Catholic University of Louvain, Louvain-la-Neuve, Belgium
anton.rodomanov@uclouvain.be

Yurii Nesterov
CORE, Catholic University of Louvain, Louvain-la-Neuve, Belgium
yurii.nesterov@uclouvain.be

The superlinear convergence of quasi-Newton methods was established as early as in 1970s, firstly by Powell [12] and Dixon [13,14] for the methods with exact line search, and then by Broyden, Dennis and Moré [15] and Dennis and Moré [16] for the methods without line search. The latter two approaches have been extended onto more general methods under various settings (see, e.g., [17–25]).

However, explicit *rates* of superlinear convergence for quasi-Newton algorithms were obtained only recently. The first results were presented in [26] for the *greedy* quasi-Newton methods. After that, in [27], the *classical* quasi-Newton methods were considered, for which the authors established certain superlinear convergence rates, depending on the problem dimension and its condition number. The analysis was based on the trace potential function, which was then augmented by the logarithm of determinant of the *inverse* Hessian approximation to extend the proof onto the general nonlinear case.

In this paper, we further improve the results of [27]. For the classical quasi-Newton methods, we obtain new convergence rate estimates, which have better dependency on the condition number of the problem. In particular, we show that the superlinear convergence rate of BFGS depends on the condition number only through the *logarithm*. As compared to the previous work, the main difference in the analysis is the choice of the potential function: now the main part is formed by the logarithm of determinant of Hessian approximation, which is then augmented by the trace of *inverse* Hessian approximation.

It is worth noting that recently, in [28], another analysis of local superlinear convergence of the classical DFP and BFGS methods was presented with the resulting rate, which is independent of the dimensionality of the problem and its condition number. However, to obtain such a rate, the authors had to make an additional assumption that the methods start from a sufficiently good initial Hessian approximation. Without this assumption, to our knowledge, their proof technique, based on the Frobenius-norm potential function, leads only to the rates, which are weaker than those in [27].

This paper is organized as follows. In Sect. 2, we introduce our notation. In Sect. 3, we study the convex Broyden class of quasi-Newton updates for approximating a self-adjoint positive definite operator. In Sect. 4, we analyze the rate of convergence of the classical quasi-Newton methods from the convex Broyden class as applied to minimizing a quadratic function. On this simple example, where the Hessian is constant, we illustrate the main ideas of our analysis. In Sect. 5, we consider the general unconstrained optimization problem. Finally, in Sect. 6, we discuss why the new superlinear convergence rates, obtained in this paper, are better than the previously known ones.

2 Notation

In what follows, $\mathbb{E}$ denotes an n-dimensional real vector space. Its dual space, composed of all linear functionals on $\mathbb{E}$, is denoted by $\mathbb{E}^*$. The value of a linear function $s \in \mathbb{E}^*$, evaluated at a point $x \in \mathbb{E}$, is denoted by $\langle s, x \rangle$.

For a smooth function $f : \mathbb{E} \to \mathbb{R}$, we denote by $\nabla f(x)$ and $\nabla^2 f(x)$ its gradient and Hessian respectively, evaluated at a point $x \in \mathbb{E}$. Note that $\nabla f(x) \in \mathbb{E}^*$, and $\nabla^2 f(x)$ is a self-adjoint linear operator from $\mathbb{E}$ to $\mathbb{E}^*$.

The partial ordering of self-adjoint linear operators is defined in the standard way. We write $A_1 \preceq A_2$ for $A_1, A_2 : \mathbb{E} \to \mathbb{E}^*$, if $\langle (A_2 - A_1)x, x \rangle \geq 0$ for all $x \in \mathbb{E}$, and $H_1 \preceq H_2$ for $H_1, H_2 : \mathbb{E}^* \to \mathbb{E}$, if $\langle s, (H_2 - H_1)s \rangle \geq 0$ for all $s \in \mathbb{E}^*$.

Any self-adjoint positive definite linear operator $A : \mathbb{E} \to \mathbb{E}^*$ induces in the spaces $\mathbb{E}$ and $\mathbb{E}^*$ the following pair of conjugate Euclidean norms:

$$\|h\|_A := \langle Ah, h \rangle^{1/2}, \quad h \in \mathbb{E}, \qquad \|s\|_A^* := \langle s, A^{-1}s \rangle^{1/2}, \quad s \in \mathbb{E}^*. \qquad (1)$$

When $A = \nabla^2 f(x)$, where $f : \mathbb{E} \to \mathbb{R}$ is a smooth function with positive definite Hessian, and $x \in \mathbb{E}$, we prefer to use notation $\|\cdot\|_x$ and $\|\cdot\|_x^*$, provided that there is no ambiguity with the reference function f.

Sometimes, in the formulas, involving products of linear operators, it is convenient to treat $x \in \mathbb{E}$ as a linear operator from $\mathbb{R}$ to $\mathbb{E}$, defined by $x\alpha = \alpha x$, and x^* as a linear operator from $\mathbb{E}^*$ to $\mathbb{R}$, defined by $x^*s = \langle s, x \rangle$. Likewise, any $s \in \mathbb{E}^*$ can be treated as a linear operator from $\mathbb{R}$ to $\mathbb{E}^*$, defined by $s\alpha = \alpha s$, and s^* as a linear operator from $\mathbb{E}$ to $\mathbb{R}$, defined by $s^*x = \langle s, x \rangle$. In this case, xx^* and ss^* are rank-one self-adjoint linear operators from $\mathbb{E}^*$ to $\mathbb{E}$ and from $\mathbb{E}^*$ to $\mathbb{E}$ respectively, acting as follows: $(xx^*)s = \langle s, x \rangle x$ and $(ss^*)x = \langle s, x \rangle s$ for $x \in \mathbb{E}$ and $s \in \mathbb{E}^*$.

Given two self-adjoint linear operators $A : \mathbb{E} \to \mathbb{E}^*$ and $H : \mathbb{E}^* \to \mathbb{E}$, we define the trace and the determinant of A with respect to H as follows: $\langle H, A \rangle := \mathrm{Tr}(HA)$, and $\mathrm{Det}(H, A) := \mathrm{Det}(HA)$. Note that HA is a linear operator from $\mathbb{E}$ to itself, and hence its trace and determinant are well-defined by the eigenvalues (they coincide with the trace and determinant of the matrix representation of HA with respect to an arbitrary chosen basis in the space $\mathbb{E}$, and the result is independent of the particular choice of the basis). In particular, if H is positive definite, then $\langle H, A \rangle$ and $\mathrm{Det}(H, A)$ are respectively the sum and the product of the eigenvalues of A relative to H^{-1}. Observe that $\langle \cdot, \cdot \rangle$ is a bilinear form, and for any $x \in \mathbb{E}$, we have $\langle Ax, x \rangle = \langle xx^*, A \rangle$. When A is invertible, we also have $\langle A^{-1}, A \rangle = n$ and $\mathrm{Det}(A^{-1}, \delta A) = \delta^n$ for any $\delta \in \mathbb{R}$. Also recall the following multiplicative formula for the determinant: $\mathrm{Det}(H, A) = \mathrm{Det}(H, G) \cdot \mathrm{Det}(G^{-1}, A)$, which is valid for any invertible linear operator $G : \mathbb{E} \to \mathbb{E}^*$. If the operator H is positive semidefinite, and $A_1 \preceq A_2$ for some self-adjoint linear operators $A_1, A_2 : \mathbb{E} \to \mathbb{E}^*$, then $\langle H, A_1 \rangle \leq \langle H, A_2 \rangle$ and $\mathrm{Det}(H, A_1) \leq \mathrm{Det}(H, A_2)$. Similarly, if A is positive semidefinite and $H_1 \preceq H_2$ for some self-adjoint linear operators $H_1, H_2 : \mathbb{E}^* \to \mathbb{E}$, then $\langle H_1, A \rangle \leq \langle H_2, A \rangle$ and $\mathrm{Det}(H_1, A) \leq \mathrm{Det}(H_2, A)$.

3 Convex Broyden Class

Let A and G be two self-adjoint positive definite linear operators from $\mathbb{E}$ to $\mathbb{E}^*$, where A is the target operator, which we want to approximate, and G is its

current approximation. The *Broyden class* of quasi-Newton updates of G with respect to A along a direction $u \in \mathbb{E} \setminus \{0\}$ is the following family of updating formulas, parameterized by a scalar $\tau \in \mathbb{R}$:

$$\begin{aligned}
\mathrm{Broyd}_\tau(A, G, u) = \phi_\tau &\left[G - \frac{Auu^*G + Guu^*A}{\langle Au, u \rangle} + \left(\frac{\langle Gu, u \rangle}{\langle Au, u \rangle} + 1 \right) \frac{Auu^*A}{\langle Au, u \rangle} \right] \\
&+ (1 - \phi_\tau) \left[G - \frac{Guu^*G}{\langle Gu, u \rangle} + \frac{Auu^*A}{\langle Au, u \rangle} \right],
\end{aligned} \tag{2}$$

where

$$\phi_\tau := \phi_\tau(A, G, u) := \frac{\tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle}}{\tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Gu, u \rangle}{\langle Au, u \rangle}}. \tag{3}$$

If the denominator in (3) is zero, we left both ϕ_τ and $\mathrm{Broyd}_\tau(A, G, u)$ undefined. For the sake of convenience, we also set $\mathrm{Broyd}_\tau(A, G, u) = G$ for $u = 0$.

In this paper, we are interested in the *convex* Broyden class, which is described by the values of $\tau \in [0, 1]$. Note that for all such τ the denominator in (3) is always positive for any $u \neq 0$, so both ϕ_τ and $\mathrm{Broyd}_\tau(A, G, u)$ are well-defined; moreover, $\phi_\tau \in [0, 1]$. For $\tau = 1$, we have $\phi_\tau = 1$, and (2) becomes the DFP update; for $\tau = 0$, we have $\phi_\tau = 0$, and (2) becomes the BFGS update.

Remark 3.1 Usually the Broyden class is defined directly in terms of the parameter ϕ. However, in the context of this paper, it is more convenient to work with τ instead of ϕ. As can be seen from (66), τ is exactly the weight of the DFP component in the updating formula for the inverse operator.

A basic property of an update from the convex Broyden class is that it preserves the bounds on the eigenvalues with respect to the target operator.

Lemma 3.1 (see [27, Lemma 2.1]) *If $\frac{1}{\xi} A \preceq G \preceq \eta A$ for some $\xi, \eta \geq 1$, then, for any $u \in \mathbb{E}$, and any $\tau \in [0, 1]$, we have $\frac{1}{\xi} A \preceq \mathrm{Broyd}_\tau(A, G, u) \preceq \eta A$.*

Consider the measure of closeness of G to A along direction $u \in \mathbb{E} \setminus \{0\}$:

$$\nu(A, G, u) := \frac{\langle (G - A)G^{-1}(G - A)u, u \rangle^{1/2}}{\langle Au, u \rangle^{1/2}} \stackrel{(1)}{=} \frac{\|(G - A)u\|_G^*}{\|u\|_A}. \tag{4}$$

Let us present two potential functions, whose improvement after one update from the convex Broyden class can be bounded from below by a certain nonnegative monotonically increasing function of ν, vanishing at zero.

First, consider the *log-det barrier*

$$V(A, G) = \ln \mathrm{Det}(A^{-1}, G). \tag{5}$$

It will be useful when $A \preceq G$. Note that in this case $V(A, G) \geq 0$.

Lemma 3.2 *Let $A, G : \mathbb{E} \to \mathbb{E}^*$ be self-adjoint positive definite linear operators, $A \preceq G \preceq \eta A$ for some $\eta \geq 1$. Then, for any $\tau \in [0, 1]$ and $u \in \mathbb{E} \setminus \{0\}$:*

$$V(A, G) - V(A, \mathrm{Broyd}_\tau(A, G, u)) \geq \ln \left(1 + (\tau \frac{1}{\eta} + 1 - \tau)\nu^2(A, G, u) \right).$$

Proof Indeed, denoting $G_+ := \mathrm{Broyd}_\tau(A, G, u)$, we obtain

$$
\begin{aligned}
V(A, G) - V(A, G_+) &\overset{(5)}{=} \ln \mathrm{Det}(G_+^{-1}, G) \\
&\overset{(67)}{=} \ln \left(\tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} \right) \\
&= \ln \left(1 + \tau \frac{\langle A(A^{-1} - G^{-1})Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle (G - A)u, u \rangle}{\langle Au, u \rangle} \right).
\end{aligned}
\tag{6}
$$

Since[1] $0 \preceq G - A \preceq (1 - \frac{1}{\eta})G$, we have

$$
(G - A)G^{-1}(G - A) \preceq \left(1 - \frac{1}{\eta} \right)(G - A) \preceq \frac{1}{1 + \frac{1}{\eta}}(G - A) \preceq G - A. \tag{7}
$$

Therefore, denoting $\nu := \nu(A, G, u)$, we can write that

$$
\frac{\langle (G-A)u, u \rangle}{\langle Au, u \rangle} \overset{(7)}{\geq} \frac{\langle (G-A)G^{-1}(G-A)u, u \rangle}{\langle Au, u \rangle} \overset{(4)}{=} \nu^2,
$$

and, since $A(A^{-1} - G^{-1})A = G - A - (G - A)G^{-1}(G - A)$, that

$$
\begin{aligned}
\frac{\langle A(A^{-1}-G^{-1})Au, u \rangle}{\langle AG^{-1}Au, u \rangle} &= \frac{\langle (G-A-(G-A)G^{-1}(G-A))u, u \rangle}{\langle AG^{-1}Au, u \rangle} \overset{(7)}{\geq} \frac{1}{\eta} \frac{\langle (G-A)G^{-1}(G-A)u, u \rangle}{\langle AG^{-1}Au, u \rangle} \\
&\geq \frac{1}{\eta} \frac{\langle (G-A)G^{-1}(G-A)u, u \rangle}{\langle Au, u \rangle} \overset{(4)}{=} \frac{1}{\eta}\nu^2.
\end{aligned}
$$

Substituting the above two inequalities into (6), we obtain the claim. □

Now consider another potential function, the *augmented log-det barrier*:

$$
\psi(G, A) := \ln \mathrm{Det}(A^{-1}, G) - \langle G^{-1}, G - A \rangle. \tag{8}
$$

As compared to the log-det barrier, this potential function is more universal since it works even if the condition $A \preceq G$ is violated. Note that the augmented log-det barrier is in fact the Bregman divergence, generated by the strictly convex function $d(A) := -\ln \mathrm{Det}(B^{-1}, A)$, defined on the set of self-adjoint positive definite linear operators from $\mathbb{E}$ to $\mathbb{E}^*$, where $B : \mathbb{E} \to \mathbb{E}^*$ is an arbitrary fixed self-adjoint positive definite linear operator. Indeed,

$$
\begin{aligned}
\psi(G, A) &= -\ln \mathrm{Det}(B^{-1}, A) + \ln \mathrm{Det}(B^{-1}, G) - \langle -G^{-1}, A - G \rangle \\
&= d(A) - d(G) - \langle \nabla d(G), A - G \rangle \geq 0.
\end{aligned}
\tag{9}
$$

Remark 3.2 The idea of combining the trace with the logarithm of determinant to form a potential function for the analysis of quasi-Newton methods can be traced back to [29]. Note also that in [27], the authors studied the evolution of $\psi(A, G)$, i.e. the Bregman divergence was centered at A instead of G.

Lemma 3.3 *For any real $\alpha \geq \beta > 0$, we have $\alpha + \frac{1}{\beta} - 1 \geq 1$, and*

$$
\alpha - \ln \beta - 1 \geq \frac{\sqrt{3}}{2 + \sqrt{3}} \ln \left(\alpha + \frac{1}{\beta} - 1 \right) \geq \frac{6}{13} \ln \left(\alpha + \frac{1}{\beta} - 1 \right). \tag{10}
$$

---

[1] This is obvious when $G - A$ is non-degenerate. The general case follows by continuity.

Proof We only need to prove the first inequality in (10) since the second one follows from it and the fact that $\frac{\sqrt{3}+2}{\sqrt{3}} = 1 + \frac{2}{\sqrt{3}} \leq 1 + \frac{7}{6} = \frac{13}{6}$ (since $2 \leq \frac{7}{2\sqrt{3}}$).

Let $\beta > 0$ be fixed, and let $\zeta_1 : (1 - \frac{1}{\beta}, +\infty) \to \mathbb{R}$ be the function, defined by $\zeta_1(\alpha) := \alpha - \frac{\sqrt{3}}{2+\sqrt{3}} \ln \left(\alpha + \frac{1}{\beta} - 1 \right)$. Note that the domain of ζ_1 includes the point $\alpha = \beta$ since $\beta \geq 2 - \frac{1}{\beta} > 1 - \frac{1}{\beta}$. Let us show that ζ_1 increases on the interval $[\beta, +\infty)$. Indeed, for any $\alpha \geq \beta$, we have

$$\zeta_1'(\alpha) = 1 - \frac{\sqrt{3}}{2+\sqrt{3}} \frac{1}{\alpha+\frac{1}{\beta}-1} \; > \; 1 - \frac{1}{\alpha+\frac{1}{\beta}-1} \; = \; \frac{\alpha+\frac{1}{\beta}-2}{\alpha+\frac{1}{\beta}-1} \; \geq \; \frac{\beta+\frac{1}{\beta}-2}{\alpha+\frac{1}{\beta}-1} \; \geq \; 0.$$

Thus, it is sufficient to prove (10) only in the case when $\alpha = \beta$. Equivalently, we need to show that the function $\zeta_2 : (0, +\infty) \to \mathbb{R}$, defined by the formula $\zeta_2(\alpha) := \alpha - \ln \alpha - 1 - \frac{\sqrt{3}}{2+\sqrt{3}} \ln \left(\alpha + \frac{1}{\alpha} - 1 \right)$, is non-negative. Differentiating, we find that, for all $\alpha > 0$, we have

$$\begin{aligned}
\zeta_2'(\alpha) &= 1 - \frac{1}{\alpha} - \frac{\sqrt{3}}{2+\sqrt{3}} \frac{1-\frac{1}{\alpha^2}}{\alpha+\frac{1}{\alpha}-1} \;=\; \left(1 - \frac{1}{\alpha}\right) \left(1 - \frac{\sqrt{3}}{2+\sqrt{3}} \frac{1+\frac{1}{\alpha}}{\alpha+\frac{1}{\alpha}-1}\right) \\
&= \left(1 - \frac{1}{\alpha}\right) \frac{\alpha+\frac{1}{\alpha}-1-(2\sqrt{3}-3)(1+\frac{1}{\alpha})}{\alpha+\frac{1}{\alpha}-1} \;=\; \left(1 - \frac{1}{\alpha}\right) \frac{\alpha-2(\sqrt{3}-1)+(\sqrt{3}-1)^2\frac{1}{\alpha}}{1+\frac{1}{\alpha}-1} \\
&= \left(1 - \frac{1}{\alpha}\right) \frac{(\sqrt{\alpha}-(\sqrt{3}-1)\frac{1}{\sqrt{\alpha}})^2}{\alpha+\frac{1}{\alpha}-1}.
\end{aligned}$$

Hence, $\zeta_2'(\alpha) \leq 0$ for $0 < \alpha \leq 1$, and $\zeta_2'(\alpha) \geq 0$ for $\alpha \geq 1$. Thus, the minimum of ζ_2 is attained at $\alpha = 1$. Consequently, $\zeta_2(\alpha) \geq \zeta_2(1) = 0$ for all $\alpha > 0$. $\square$

It turns out that, up to some constants, the improvement in the augmented log-det barrier can be bounded from below by exactly the same logarithmic function of ν, which was used for the simple log-det barrier.

Lemma 3.4 *Let $A, G : \mathbb{E} \to \mathbb{E}^*$ be self-adjoint positive definite linear operators, $\frac{1}{\xi} A \preceq G \preceq \eta A$ for some $\xi, \eta \geq 1$. Then, for any $\tau \in [0, 1]$ and $u \in \mathbb{E} \setminus \{0\}$:*

$$\psi(G, A) - \psi(\mathrm{Broyd}_\tau(A, G, u), A) \geq \tfrac{6}{13} \ln \left(1 + (\tau \tfrac{1}{\xi\eta} + 1 - \tau)\nu^2(A, G, u) \right).$$

Proof Indeed, denoting $G_+ := \mathrm{Broyd}_\tau(A, G, u)$, we obtain

$$\langle G^{-1} - G_+^{-1}, A \rangle \overset{(66)}{=} \tau \left[\frac{\langle AG^{-1}AG^{-1}Au, u \rangle}{\langle AG^{-1}Au, u \rangle} - 1 \right] + (1 - \tau) \left[\frac{\langle AG^{-1}Au, u \rangle}{\langle Au, u \rangle} - 1 \right],$$

and

$$\mathrm{Det}(G_+^{-1}, G) \overset{(67)}{=} \tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Au, u \rangle}{\langle Gu, u \rangle}.$$

Thus,

$$\begin{aligned}
\psi(G, A) - \psi(G_+, A) &\overset{(8)}{=} \langle G^{-1} - G_+^{-1}, A \rangle + \ln \mathrm{Det}(G_+^{-1}, G) \\
&= \tau\alpha_1 + (1 - \tau)\alpha_0 + \ln(\tau\beta_1^{-1} + (1 - \tau)\beta_0^{-1}) - 1 \;=\; \alpha - \ln\beta - 1,
\end{aligned} \tag{11}$$

where we denote $\alpha_1 := \frac{\langle AG^{-1}AG^{-1}Au, u\rangle}{\langle AG^{-1}Au, u\rangle}$, $\beta_1 := \frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle}$, $\alpha_0 := \frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle}$, $\beta_0 := \frac{\langle Au, u\rangle}{\langle Gu, u\rangle}$, $\alpha := \tau\alpha_1 + (1-\tau)\alpha_0$, $\beta := (\tau\beta_1^{-1} + (1-\tau)\beta_0^{-1})^{-1}$. Note that $\alpha_1 \geq \beta_1$ and $\alpha_0 \geq \beta_0$ by the Cauchy-Schwartz inequality. At the same time, $\tau\beta_1 + (1-\tau)\beta_2 \geq \beta$ by the convexity of the inverse function $t \mapsto t^{-1}$. Hence, we can apply Lemma 3.3 to estimate (11) from below. Note that

$$
\begin{aligned}
\alpha + \tfrac{1}{\beta} - 1 &= \tau\frac{\langle(A+AG^{-1}AG^{-1}A)u, u\rangle}{\langle AG^{-1}Au, u\rangle} + (1-\tau)\frac{\langle(G+A)u, u\rangle}{\langle Au, u\rangle} - 1 \\
&= 1 + \tau\frac{\langle(G-A)G^{-1}AG^{-1}(G-A)\rangle}{\langle AG^{-1}Au, u\rangle} + (1-\tau)\frac{\langle(G-A)G^{-1}(G-A)u, u\rangle}{\langle Au, u\rangle} \\
&\geq 1 + (\tau\tfrac{1}{\xi\eta} + 1 - \tau)\frac{\langle(G-A)G^{-1}(G-A)u, u\rangle}{\langle Au, u\rangle} \\
&\overset{(4)}{=} 1 + (\tau\tfrac{1}{\xi\eta} + 1 - \tau)\nu^2(A, G, u). \qquad \square
\end{aligned}
$$

The measure $\nu(A, G, u)$, defined in (4), is the ratio of the norm of $(G-A)u$, measured with respect to G, and the norm of u, measured with respect to A. It is important that we can change the corresponding metrics to G_+ and G respectively by paying only with the minimal eigenvalue of G relative to A.

Lemma 3.5 *Let $A, G : \mathbb{E} \to \mathbb{E}^*$ be self-adjoint positive definite linear operators such that $\frac{1}{\xi}A \preceq G$ for some $\xi > 0$. Then, for any $\tau \in [0, 1]$, any $u \in \mathbb{E} \setminus \{0\}$, and $G_+ := \mathrm{Broyd}_\tau(A, G, u)$, we have*

$$
\nu^2(A, G, u) \geq \tfrac{1}{1+\xi}\frac{\langle(G-A)G_+^{-1}(G-A)u, u\rangle}{\langle Gu, u\rangle}.
$$

Proof From (66), it is easy to see that $G_+^{-1}Au = u$. Hence,

$$
\begin{aligned}
\frac{\langle(G-A)G_+^{-1}(G-A)u, u\rangle}{\langle Gu, u\rangle} &= \frac{\langle GG_+^{-1}Gu, u\rangle}{\langle Gu, u\rangle} + \frac{\langle Au, G_+^{-1}Au\rangle}{\langle Gu, u\rangle} - 2\frac{\langle Gu, G_+^{-1}Au\rangle}{\langle Gu, u\rangle} \\
&= \frac{\langle GG_+^{-1}Gu, u\rangle}{\langle Gu, u\rangle} + \frac{\langle Au, u\rangle}{\langle Gu, u\rangle} - 2.
\end{aligned} \tag{12}
$$

Since $1 - t \leq \frac{1}{t} - 1$ for all $t > 0$, we further have

$$
\begin{aligned}
\frac{\langle GG_+^{-1}Gu, u\rangle}{\langle Gu, u\rangle} \overset{(66)}{=} &\; \tau\left[1 - \frac{\langle Au, u\rangle^2}{\langle Gu, u\rangle\langle AG^{-1}Au, u\rangle} + \frac{\langle Gu, u\rangle}{\langle Au, u\rangle}\right] \\
&+ (1-\tau)\left[\left(\frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} + 1\right)\frac{\langle Gu, u\rangle}{\langle Au, u\rangle} - 1\right] \\
\leq &\; \left(\frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} + 1\right)\frac{\langle Gu, u\rangle}{\langle Au, u\rangle} - 1.
\end{aligned} \tag{13}
$$

Denote $\nu := \nu(A, G, u)$. Then,

$$
\nu^2 \overset{(4)}{=} \frac{\langle(G-A)G^{-1}(G-A)u, u\rangle}{\langle Au, u\rangle} = \frac{\langle Gu, u\rangle}{\langle Au, u\rangle} + \frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} - 2. \tag{14}
$$

Consequently,

$$
\begin{aligned}
(1+\xi)\nu^2 &\geq \left(\frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} + 1\right)\nu^2 \\
&\overset{(14)}{=} \left(\frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} + 1\right)\frac{\langle Gu, u\rangle}{\langle Au, u\rangle} + \frac{\langle AG^{-1}Au, u\rangle^2}{\langle Au, u\rangle^2} - \frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} - 2 \\
&\overset{(13)}{\geq} \frac{\langle GG_+^{-1}Gu, u\rangle}{\langle Au, u\rangle} + \frac{\langle AG^{-1}Au, u\rangle^2}{\langle Au, u\rangle} - \frac{\langle AG^{-1}Au, u\rangle}{\langle Au, u\rangle} - 1.
\end{aligned} \tag{15}
$$

Thus,

$$(1+\xi)\nu^2 - \frac{\langle (G-A)G_+^{-1}(G-A)u,u\rangle}{\langle Gu,u\rangle} \overset{(12)}{=} (1+\xi)\nu^2 - \frac{\langle GG_+^{-1}Gu,u\rangle}{\langle Gu,u\rangle} - \frac{\langle Au,u\rangle}{\langle Gu,u\rangle} + 2$$

$$\overset{(15)}{\geq} \frac{\langle AG^{-1}Au,u\rangle^2}{\langle Au,u\rangle^2} - \frac{\langle AG^{-1}Au,u\rangle}{\langle Au,u\rangle} - \frac{\langle Au,u\rangle}{\langle Gu,u\rangle} + 1$$

$$\geq \frac{\langle AG^{-1}Au,u\rangle^2}{\langle Au,u\rangle^2} - 2\frac{\langle AG^{-1}Au,u\rangle}{\langle Au,u\rangle} + 1 \geq 0,$$

where we have used the Cauchy–Schwartz inequality $\frac{\langle Au,u\rangle}{\langle Gu,u\rangle} \leq \frac{\langle AG^{-1}Au,u\rangle}{\langle Au,u\rangle}$. $\quad\square$

4 Unconstrained Quadratic Minimization

Let us study the convergence properties of the classical quasi-Newton methods from the convex Broyden class, as applied to minimizing the quadratic function

$$f(x) := \tfrac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle, \tag{16}$$

where $A : \mathbb{E} \to \mathbb{E}^*$ is a self-adjoint positive definite linear operator, and $b \in \mathbb{E}^*$.

Let $B : \mathbb{E} \to \mathbb{E}^*$ be a fixed self-adjoint positive definite linear operator, and let $\mu, L > 0$ be such that

$$\mu B \preceq A \preceq LB. \tag{17}$$

Thus, μ is the *strong convexity* parameter of f, and L is the constant of *Lipschitz continuity* of the gradient of f, both measured relative to B.

Consider the following standard quasi-Newton process for minimizing (16):

$$
\boxed{
\begin{array}{l}
\textbf{Initialization: } \text{Choose } x_0 \in \mathbb{E}. \text{ Set } G_0 = LB. \\[4pt]
\textbf{For } k \geq 0 \textbf{ iterate:} \\[4pt]
\text{1. Update } x_{k+1} = x_k - G_k^{-1}\nabla f(x_k). \\[4pt]
\text{2. Set } u_k = x_{k+1} - x_k \text{ and choose } \tau_k \in [0,1]. \\[4pt]
\text{3. Compute } G_{k+1} = \text{Broyd}_{\tau_k}(A, G_k, u_k).
\end{array}
}
\tag{18}
$$

For measuring its rate of convergence, we use the norm of the gradient, taken with respect to the Hessian:

$$\lambda_k := \|\nabla f(x_k)\|_A^* \overset{(1)}{=} \langle \nabla f(x_k), A^{-1}\nabla f(x_k)\rangle^{1/2}.$$

It is known that the process (18) has at least a linear convergence rate of the standard gradient method:

Theorem 4.1 (see [27, Theorem 3.1]) *In scheme* (18), *for all* $k \geq 0$:

$$A \preceq G_k \preceq \tfrac{L}{\mu}A, \qquad \lambda_k \leq \left(1 - \tfrac{\mu}{L}\right)^k \lambda_0. \tag{19}$$

Let us establish the superlinear convergence. According to (19), for the quadratic function, we have $A \preceq G_k$ for all $k \geq 0$. Therefore, in our analysis, we can use both potential functions: the log-det barrier and the augmented log-det barrier. Let us consider both options. We start with the first one.

Theorem 4.2 *In scheme* (18), *for all $k \geq 1$, we have*

$$\lambda_k \leq \left[\frac{2}{\prod_{i=0}^{k-1}(\tau_i \frac{\mu}{L}+1-\tau_i)^{1/k}} \left(e^{\frac{n}{k} \ln \frac{L}{\mu}} - 1 \right) \right]^{k/2} \sqrt{\frac{L}{\mu}} \cdot \lambda_0. \tag{20}$$

Proof Without loss of generality, we can assume that $u_i \neq 0$ for all $0 \leq i \leq k$. Denote $V_i := V(A, G_i)$, $\nu_i := \nu(A, G_i, u_i)$, $p_i := \tau_i \frac{\mu}{L}+1-\tau_i$, $g_i := \|\nabla f(x_i)\|_{G_i}^*$ for any $0 \leq i \leq k$. By Lemma 3.2 and (19), for all $0 \leq i \leq k-1$, we have $\ln(1 + p_i \nu_i^2) \leq V_i - V_{i+1}$. Summing up, we obtain

$$
\begin{aligned}
\sum_{i=0}^{k-1} \ln(1 + p_k \nu_k^2) &\leq V_0 - V_k \overset{(19)}{\leq} V_0 \overset{(18)}{=} V(A, LB) \\
&\overset{(5)}{=} \ln \operatorname{Det}(A^{-1}, LB) \overset{(17)}{\leq} \ln \operatorname{Det}(\tfrac{1}{\mu}B^{-1}, LB) = n \ln \tfrac{L}{\mu}.
\end{aligned}
\tag{21}
$$

Hence, by the convexity of function $t \mapsto \ln(1 + e^t)$, we get

$$
\begin{aligned}
\tfrac{n}{k} \ln \tfrac{L}{\mu} \overset{(21)}{\geq} \tfrac{1}{k} \sum_{i=0}^{k-1} \ln(1 + p_i \nu_i^2) &= \tfrac{1}{k} \sum_{i=0}^{k-1} \ln(1 + e^{\ln(p_i \nu_i^2)}) \\
&\geq \ln \left(1 + e^{\frac{1}{k} \sum_{i=0}^{k-1} \ln(p_i \nu_i^2)} \right) = \ln \left(1 + \left[\prod_{i=0}^{k-1} p_i \nu_i^2 \right]^{1/k} \right).
\end{aligned}
\tag{22}
$$

But, for all $0 \leq i \leq k-1$, we have $\nu_i^2 \geq \frac{1}{2} \frac{\langle (G_i-A)G_{i+1}^{-1}(G_i-A)u_i, u_i \rangle}{\langle G_i u_i, u_i \rangle} = \frac{1}{2} \frac{g_{i+1}^2}{g_i^2}$ by Lemma 3.5, (19), and since $G_i u_i = -\nabla f(x_i)$, $Au_i = \nabla f(x_{i+1}) - \nabla f(x_i)$. Hence, $\prod_{i=0}^{k-1} \nu_i^2 \geq \frac{1}{2^k} \frac{g_k^2}{g_0^2}$, and so $\frac{n}{k} \ln \frac{L}{\mu} \overset{(22)}{\geq} \ln \left(1 + \frac{1}{2} \left[\prod_{i=0}^{k-1} p_i \right]^{1/k} \left[\frac{g_k}{g_0} \right]^{2/k} \right)$.

Rearranging, we obtain $g_k \leq \left[\frac{2}{\prod_{i=0}^{k-1} p_i^{1/k}} (e^{\frac{n}{k} \ln \frac{L}{\mu}} - 1) \right]^{k/2} g_0$. It remains to note that $\lambda_k \leq \sqrt{\frac{L}{\mu}} \cdot g_k$ and $g_0 \leq \lambda_0$ in view of (19). $\qquad \square$

Remark 4.1 As can be seen from (21), the factor $n \ln \frac{L}{\mu}$ in (20) can be improved up to $\ln \operatorname{Det}(A^{-1}, LB) = \sum_{i=1}^{n} \ln \frac{L}{\lambda_i}$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A relative to B. This improved factor can be significantly smaller than the original one if the majority of the eigenvalues λ_i are much larger than μ.

Let us briefly present another approach, which is based on the *augmented* log-det barrier. The resulting efficiency estimate will be the same as in Theorem 4.2 up to a slightly worse absolute constant under the exponent. However, this proof can be extended onto general nonlinear functions.

Theorem 4.3 *In scheme* (18), *for all* $k \geq 1$, *we have*

$$\lambda_k \leq \left[\frac{2}{\prod_{i=0}^{k-1}(\tau_i \frac{\mu}{L}+1-\tau_i)^{1/k}} \left(e^{\frac{13}{6}\frac{n}{k}\ln\frac{L}{\mu}} - 1 \right) \right]^{k/2} \sqrt{\frac{L}{\mu}} \cdot \lambda_0.$$

Proof Without loss of generality, we can assume that $u_i \neq 0$ for all $0 \leq i \leq k$. Denote $\psi_i := \psi(G_i, A)$, $\nu_i := \nu(A, G_i, u_i)$, $p_i = \tau_i \frac{\mu}{L} + 1 - \tau_i$, $g_i := \|\nabla f(x_i)\|_{G_i}^*$ for all $0 \leq i \leq k$. By Lemma 3.4 and (19), for all $0 \leq i \leq k-1$, we have $\frac{6}{13}\ln(1 + p_i \nu_i^2) \leq \psi_i - \psi_{i+1}$. Hence,

$$\frac{6}{13}\sum_{i=0}^{k-1}\ln(1 + p_i\nu_i^2) \;\leq\; \psi_0 - \psi_k \;\overset{(9)}{\leq}\; \psi_0 \;\overset{(18)}{=}\; \psi(LB, A)$$

$$\overset{(8)}{=}\; \ln\mathrm{Det}(A^{-1}, LB) - \langle \tfrac{1}{L}B^{-1}, LB - A \rangle \;\overset{(17)}{\leq}\; n\ln\frac{L}{\mu}, \tag{23}$$

and we can continue exactly as in the proof of Theorem 4.2. $\square$

5 Minimization of General Functions

In this section, we consider the general unconstrained minimization problem:

$$\min_{x\in\mathbb{E}} f(x), \tag{24}$$

where $f : \mathbb{E} \to \mathbb{R}$ is a twice continuously differentiable function with positive definite second derivative. Our goal is to study the convergence properties of the following standard quasi-Newton scheme for solving (24):

$$\begin{array}{|l|} \hline \\[-0.5em]
\textbf{Initialization:} \text{ Choose } x_0 \in \mathbb{E}. \text{ Set } G_0 = LB. \\[0.8em]
\textbf{For } k \geq 0 \textbf{ iterate:} \\[0.8em]
\text{1. Update } x_{k+1} = x_k - G_k^{-1}\nabla f(x_k). \\[0.8em]
\text{2. Set } u_k = x_{k+1} - x_k \text{ and choose } \tau_k \in [0,1]. \\[0.8em]
\text{3. Denote } J_k = \int_0^1 \nabla^2 f(x_k + tu_k)dt. \\[0.8em]
\text{4. Set } G_{k+1} = \mathrm{Broyd}_{\tau_k}(J_k, G_k, u_k). \\[0.5em]
\hline \end{array} \tag{25}$$

Here $B : \mathbb{E} \to \mathbb{E}^*$ is a self-adjoint positive definite linear operator, and L is a positive constant, which together define the initial Hessian approximation G_0.

We assume that there exist constants $\mu > 0$ and $M \geq 0$, such that

$$\mu B \preceq \nabla^2 f(x) \preceq LB, \tag{26}$$

$$\nabla^2 f(y) - \nabla^2 f(x) \preceq M\|y - x\|_z \nabla^2 f(w) \tag{27}$$

for all $x, y, z, w \in \mathbb{E}$. The first assumption (26) specifies that, relative to the operator B, the objective function f is μ-*strongly convex* and its gradient is *L-Lipschitz continuous*. The second assumption (27) means that f is M-*strongly self-concordant*. This assumption was recently introduced in [26] as a convenient affine-invariant alternative to the standard assumption of the Lipschitz second derivative, and is satisfied at least for any strongly convex function with Lipschitz continuous Hessian (see [26, Example 4.1]). The main facts, which we use about strongly self-concordant functions, are summarized in the following lemma (see [26, Lemma 4.1]):

Lemma 5.1 *For any $x, y \in \mathbb{E}$, $J := \int_0^1 \nabla^2 f(x + t(y - x))dt$, $r := \|y - x\|_x$:*

$$\left(1 + \tfrac{Mr}{2}\right)^{-1} \nabla^2 f(x) \preceq J \preceq \left(1 + \tfrac{Mr}{2}\right) \nabla^2 f(x), \tag{28}$$

$$\left(1 + \tfrac{Mr}{2}\right)^{-1} \nabla^2 f(y) \preceq J \preceq \left(1 + \tfrac{Mr}{2}\right) \nabla^2 f(y). \tag{29}$$

Note that for a quadratic function, we have $M = 0$.

For measuring the convergence rate of (25), we use the local gradient norm:

$$\lambda_k := \|\nabla f(x_k)\|_{x_k}^* \overset{(1)}{=} \langle \nabla f(x_k), \nabla^2 f(x_k)^{-1} \nabla f(x_k) \rangle^{1/2}. \tag{30}$$

The local convergence analysis of the scheme (25) is, in general, the same as the corresponding analysis in the quadratic case. However, it is much more technical due to the fact that, in the nonlinear case, the Hessian is no longer constant. This causes a few problems.

First, there are now several different ways how one can treat the Hessian approximation G_k. One can view it as an approximation to the Hessian $\nabla^2 f(x_k)$ at the current iterate x_k, to the Hessian $\nabla^2 f(x^*)$ at the minimizer x^*, to the integral Hessian J_k etc. Of course, locally, due to strong self-concordancy, all these variants are equivalent since the corresponding Hessians are close to each other. Nevertheless, from the viewpoint of technical simplicity of the analysis, some options are slightly more preferable than others. We find it to be the most convenient to always think of G_k as an approximation to the integral Hessian J_k.

The second issue is as follows. Suppose we already know what is the connection between our current Hessian approximation G_k and the actual integral Hessian J_k, e.g., in terms of the relative eigenvalues and the value of the augmented log-det barrier potential function (8). Naturally, we want to know how these quantities change after we update G_k into G_{k+1} at Step 4 of the scheme (25). For this, we apply Lemma 3.1 and Lemma 3.4 respectively. However, the problem is that both of these lemmas will provide us only with the information on the connection between the update result G_{k+1} and the *current* integral Hessian J_k (which was used for performing the update), not the next one J_{k+1}. Therefore, we need to additionally take into account the errors, resulting from approximating J_{k+1} by J_k.

For estimating the errors, which accumulate as a result of approximating one Hessian by another, it is convenient to introduce the following quantities[2]:

$$r_k := \|u_k\|_{x_k}, \qquad \xi_k := e^{M \sum_{i=0}^{k-1} r_i} \quad (\geq 1), \qquad k \geq 0. \tag{31}$$

Remark 5.1 The general framework of our analysis is the same as in the previous paper [27]. The main difference is that now another potential function is used for establishing the rate of superlinear convergence (Lemma 5.4). However, in order to properly incorporate the new potential function into the analysis, many parts in the proof had to be appropriately modified, most notably the part, related to estimating the region of local convergence. In any case, the analysis, presented below, is fully self-contained, and does not require the reader first go through [27].

We analyze the method (25) in several steps. The first step is to establish the bounds on the relative eigenvalues of the Hessian approximations with respect to the corresponding Hessians.

Lemma 5.2 *For all $k \geq 0$, we have*

$$\tfrac{1}{\xi_k} \nabla^2 f(x_k) \preceq G_k \preceq \xi_k \tfrac{L}{\mu} \nabla^2 f(x_k), \tag{32}$$

$$\tfrac{1}{\xi_{k+1}} J_k \preceq G_k \preceq \xi_{k+1} \tfrac{L}{\mu} J_k. \tag{33}$$

Proof For $k = 0$, (32) follows from (26) and the fact that $G_0 = LB$ and $\xi_0 = 1$. Now suppose that $k \geq 0$, and that (32) has already been proved for all indices up to k. Then, applying Lemma 5.1 to (32), we obtain

$$\frac{1}{\xi_k \left(1 + \frac{Mr_k}{2}\right)} J_k \preceq G_k \preceq \left(1 + \frac{Mr_k}{2}\right) \xi_k \tfrac{L}{\mu} J_k. \tag{34}$$

Since $(1 + \frac{Mr_k}{2})\xi_k \leq \xi_{k+1}$ by (31), this proves (33) for the index k. Applying Lemma 3.1 to (34), we get $\frac{1}{\xi_k(1+\frac{Mr_k}{2})} J_k \preceq G_{k+1} \preceq (1 + \frac{Mr_k}{2})\xi_k \tfrac{L}{\mu} J_k$, and so

$$G_{k+1} \overset{(29)}{\preceq} \left(1 + \tfrac{Mr_k}{2}\right)^2 \xi_k \tfrac{L}{\mu} \nabla^2 f(x_{k+1}) \overset{(31)}{\preceq} \xi_{k+1} \tfrac{L}{\mu} \nabla^2 f(x_{k+1}),$$

$$G_{k+1} \overset{(29)}{\succeq} \frac{1}{\left(1 + \frac{Mr_k}{2}\right)^2 \xi_k} \nabla^2 f(x_{k+1}) \overset{(31)}{\succeq} \frac{1}{\xi_{k+1}} \nabla^2 f(x_{k+1}).$$

This proves (32) for the index $k + 1$, and we can continue by induction. □

Corollary 5.1 *For all $k \geq 0$, we have*

$$r_k \leq \xi_k \lambda_k. \tag{35}$$

---

[2] We follow the standard convention that the sum over the empty set is defined as 0, so $\xi_0 = 1$. Similarly, the product over the empty set is defined as 1.

Proof Indeed,

$$r_k \overset{(31)}{=} \|u_k\|_{x_k} \overset{(25)}{=} \langle \nabla f(x_k), G_k^{-1} \nabla^2 f(x_k) G_k^{-1} \nabla f(x_k) \rangle^{1/2}$$
$$\overset{(32)}{\leq} \xi_k \langle \nabla f(x_k), \nabla^2 f(x_k)^{-1} \nabla f(x_k) \rangle^{1/2} \overset{(30)}{=} \xi_k \lambda_k. \qquad \square$$

The second step in our analysis is to establish a preliminary version of the linear convergence theorem for the scheme (25).

Lemma 5.3 *For all $k \geq 0$, we have*

$$\lambda_k \leq \sqrt{\xi_k} \lambda_0 \prod_{i=0}^{k-1} q_i, \qquad (36)$$

where

$$q_i := \max \left\{ 1 - \frac{\mu}{\xi_{i+1}L}, \xi_{i+1} - 1 \right\}. \qquad (37)$$

Proof Let $k, i \geq 0$ be arbitrary. By Taylor's formula, we have

$$\nabla f(x_{i+1}) \overset{(25)}{=} \nabla f(x_i) + J_i u_i \overset{(25)}{=} J_i(J_i^{-1} - G_i^{-1}) \nabla f(x_i). \qquad (38)$$

Hence,

$$\lambda_{i+1} \overset{(30)}{=} \langle \nabla f(x_{i+1}), \nabla^2 f(x_{i+1})^{-1} \nabla f(x_{i+1}) \rangle^{1/2}$$
$$\overset{(29)}{\leq} \sqrt{1 + \frac{Mr_i}{2}} \langle \nabla f(x_{i+1}), J_i^{-1} \nabla f(x_{i+1}) \rangle^{1/2} \qquad (39)$$
$$\overset{(38)}{=} \sqrt{1 + \frac{Mr_i}{2}} \langle \nabla f(x_i), (J_i^{-1} - G_i^{-1}) J_i (J_i^{-1} - G_i^{-1}) \nabla f(x_i) \rangle^{1/2}.$$

Note that $-(\xi_{i+1} - 1) J_i^{-1} \overset{(33)}{\preceq} J_i^{-1} - G_i^{-1} \overset{(33)}{\preceq} \left(1 - \frac{\mu}{\xi_{i+1}L} \right) J_i^{-1}$. Therefore,

$$(J_i^{-1} - G_i^{-1}) J_i (J_i^{-1} - G_i^{-1}) \overset{(37)}{\preceq} q_i^2 J_i^{-1} \overset{(28)}{\preceq} q_i^2 \left(1 + \frac{Mr_i}{2} \right) \nabla^2 f(x_i)^{-1}.$$

Thus, $\lambda_{i+1} \leq \left(1 + \frac{Mr_i}{2} \right) q_i \lambda_i$ in view of (39) and (30). Consequently,

$$\lambda_k \leq \lambda_0 \prod_{i=0}^{k-1} \left(1 + \frac{Mr_i}{2} \right) q_i \leq \lambda_0 \prod_{i=0}^{k-1} e^{\frac{Mr_i}{2}} q_i \overset{(31)}{=} \sqrt{\xi_k} \lambda_0 \prod_{i=0}^{k-1} q_i. \qquad \square$$

Next, we establish a preliminary version of the theorem on superlinear convergence of the scheme (25). The proof uses the augmented log-det barrier potential function and is essentially a generalization of the corresponding proof of Theorem 4.3.

Lemma 5.4 *For all $k \geq 1$, we have*

$$\lambda_k \leq \left[\frac{1 + \xi_k}{\prod_{i=0}^{k-1} (\tau_i \frac{\mu}{\xi_{i+1}^2 L} + 1 - \tau_i)^{1/k}} \left(e^{\frac{13}{6} \frac{n}{k} \ln\left(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu} \right)} - 1 \right) \right]^{k/2} \sqrt{\xi_k \frac{L}{\mu}} \cdot \lambda_0. \qquad (40)$$

Proof Without loss of generality, assume that $u_i \neq 0$ for all $0 \leq i \leq k$. Denote $\psi_i := \psi(G_i, J_i)$, $\tilde{\psi}_{i+1} := \psi(G_{i+1}, J_i)$, $\nu_i := \nu(J_i, G_i, u_i)$, $p_i := \tau_i \frac{\mu}{\xi_{i+1}^2 L} + 1 - \tau_i$, and $g_i := \|\nabla f(x_i)\|_{G_i}^*$ for any $0 \leq i \leq k$.

Let $0 \leq i \leq k - 1$ be arbitrary. By Lemma 3.4 and (33), we have

$$\tfrac{6}{13} \ln\left(1 + p_i \nu_i^2\right) \leq \psi_i - \tilde{\psi}_{i+1} \;=\; \psi_i - \psi_{i+1} + \Delta_i, \tag{41}$$

where

$$\Delta_i := \psi_{i+1} - \tilde{\psi}_{i+1} \overset{(8)}{=} \langle G_{i+1}^{-1}, J_{i+1} - J_i \rangle + \ln \mathrm{Det}(J_{i+1}^{-1}, J_i). \tag{42}$$

Note that $J_i \succeq (1 + \frac{Mr_i}{2})^{-1} \nabla^2 f(x_{i+1}) \succeq (1 + \frac{Mr_i}{2})^{-1}(1 + \frac{Mr_{i+1}}{2})^{-1} J_{i+1}$ in view of (29) and (28). In particular, $J_i \succeq e^{-\frac{M}{2}(r_i + r_{i+1})} J_{i+1} \succeq (1 - \frac{M}{2}(r_i + r_{i+1})) J_{i+1}$. Therefore, $J_{i+1} - J_i \preceq \frac{M}{2}(r_i + r_{i+1}) J_{i+1}$, and so

$$
\begin{aligned}
\sum_{i=0}^{k-1} \langle G_{i+1}^{-1}, J_{i+1} - J_i \rangle
&\leq \frac{M}{2} \sum_{i=0}^{k-1} (r_i + r_{i+1}) \langle G_{i+1}^{-1}, J_{i+1} \rangle \\
&\overset{(33)}{\leq} n\frac{M}{2} \sum_{i=0}^{k-1} \xi_{i+2}(r_i + r_{i+1}) \overset{(31)}{\leq} n\xi_{k+1} \frac{M}{2} \sum_{i=0}^{k-1} (r_i + r_{i+1}) \\
&\leq n\xi_{k+1} M \sum_{i=0}^{k} r_i \overset{(31)}{=} n\xi_{k+1} \ln \xi_{k+1}.
\end{aligned}
$$

Consequently,

$$\sum_{i=0}^{k-1} \Delta_i \overset{(42)}{\leq} n\xi_{k+1} \ln \xi_{k+1} + \ln \mathrm{Det}(J_k^{-1}, J_0). \tag{43}$$

Summing up (41), we thus obtain

$$
\begin{aligned}
\tfrac{6}{13} \sum_{i=0}^{k-1} \ln(1 + p_i \nu_i^2) &\leq \psi_0 - \psi_k + \sum_{i=0}^{k-1} \Delta_i \overset{(9)}{\leq} \psi_0 + \sum_{i=0}^{k-1} \Delta_i \\
&\overset{(8)}{=} \ln \mathrm{Det}(J_0^{-1}, LB) - \langle \tfrac{1}{L} B^{-1}, LB - J_0 \rangle + \sum_{i=0}^{k-1} \Delta_i \\
&\overset{(43)}{\leq} \ln \mathrm{Det}(J_k^{-1}, LB) - \langle \tfrac{1}{L} B^{-1}, LB - J_0 \rangle + n\xi_{k+1} \ln \xi_{k+1} \\
&\overset{(26)}{\leq} n \ln \tfrac{L}{\mu} + n\xi_{k+1} \ln \xi_{k+1} \;=\; n \ln\left(\xi_{k+1}^{\xi_{k+1}} \tfrac{L}{\mu}\right).
\end{aligned}
$$

By the convexity of function $t \mapsto \ln(1 + e^t)$, it follows that

$$
\begin{aligned}
\tfrac{13}{6} \tfrac{n}{k} \ln\left(\xi_{k+1}^{\xi_{k+1}} \tfrac{L}{\mu}\right) &\geq \tfrac{1}{k} \sum_{i=0}^{k-1} \ln(1 + p_i \nu_i^2) = \tfrac{1}{k} \sum_{i=0}^{k-1} \ln(1 + e^{\ln(p_i \nu_i^2)}) \\
&\geq \ln\left(1 + e^{\frac{1}{k} \sum_{i=0}^{k-1} \ln(p_i \nu_i^2)}\right) = \ln\left(1 + \left[\prod_{i=0}^{k-1} p_i \nu_i^2\right]^{1/k}\right).
\end{aligned} \tag{44}
$$

At the same time, $\nu_i^2 \geq \frac{1}{1 + \xi_{i+1}} \frac{\langle (G_i - J_i) G_{i+1}^{-1} (G_i - J_i) u_i, u_i \rangle}{\langle G_i u_i, u_i \rangle} = \frac{1}{1 + \xi_{i+1}} \frac{g_{i+1}^2}{g_i^2}$ in view of Lemma 3.5, (33) and since $G_i u_i = -\nabla f(x_i)$, $J_i u_i = \nabla f(x_{i+1}) - \nabla f(x_i)$.

Hence, we can write $\prod_{i=0}^{k-1} \nu_i^2 \geq \frac{g_k^2}{g_0^2} \prod_{i=0}^{k-1} \frac{1}{1+\xi_{i+1}} \overset{(31)}{\geq} \frac{1}{(1+\xi_k)^k} \frac{g_k^2}{g_0^2}$. Consequently, $\frac{13}{6} \frac{n}{k} \ln(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu}) \overset{(44)}{\geq} \ln\left(1 + \frac{\prod_{i=0}^{k-1} p_i^{1/k}}{1+\xi_k} \left[\frac{g_k}{g_0}\right]^{2/k}\right)$. Rearranging, we obtain that $g_k \leq \left[\frac{1+\xi_k}{\prod_{i=0}^{k-1} p_i^{1/k}} (e^{\frac{13}{6} \frac{n}{k} \ln(\xi_{k+1}^{\xi_{k+1}} \frac{L}{\mu})} - 1)\right]^{k/2} g_0$. But $\lambda_k \leq \sqrt{\xi_k \frac{L}{\mu}} \cdot g_k$ by (32), and $g_0 \leq \lambda_0$ in view of (26) and the fact that $G_0 = LB$. $\qquad\square$

In the quadratic case ($M = 0$), we have $\xi_k \equiv 1$ (see (31)), and Lemmas 5.2 and 5.3 reduce to the already known Theorem 4.1, and Lemma 5.4 reduces to the already known Theorem 4.2. In the general case, the quantities ξ_k can grow with iterations. However, as we will see in a moment, by requiring the initial point x_0 in the scheme (25) to be sufficiently close to the solution, we can still ensure that ξ_k stay *uniformly bounded* by a sufficiently small absolute constant. This allows us to recover all the main results of the quadratic case.

To write down the region of local convergence of (25), we need to introduce one more quantity, related to the starting moment of superlinear convergence[3]:

$$K_0 := \left\lceil \frac{1}{\tau \frac{4\mu}{9L} + 1 - \tau} 8n \ln \frac{2L}{\mu} \right\rceil, \qquad \tau := \sup_{k \geq 0} \tau_k \quad (\leq 1). \tag{45}$$

For DFP ($\tau_k \equiv 1$) and BFGS ($\tau_k \equiv 0$), we have respectively

$$K_0^{\text{DFP}} = \left\lceil \frac{18nL}{\mu} \ln \frac{2L}{\mu} \right\rceil, \qquad K_0^{\text{BFGS}} = \left\lceil 8n \ln \frac{2L}{\mu} \right\rceil. \tag{46}$$

Now we are ready to prove the main result of this section.

Theorem 5.1 *Suppose that, in scheme* (25)*, we have*

$$M\lambda_0 \leq \frac{\ln \frac{3}{2}}{\left(\frac{3}{2}\right)^{\frac{3}{2}}} \max\left\{\frac{\mu}{2L}, \frac{1}{K_0+9}\right\}. \tag{47}$$

Then, for all $k \geq 0$,

$$\frac{2}{3} \nabla^2 f(x_k) \preceq G_k \preceq \frac{3L}{2\mu} \nabla^2 f(x_k), \tag{48}$$

$$\lambda_k \leq \left(1 - \frac{\mu}{2L}\right)^k \sqrt{\frac{3}{2}} \cdot \lambda_0, \tag{49}$$

and, for all $k \geq 1$,

$$\lambda_k \leq \left[\frac{5}{2 \prod_{i=0}^{k-1} (\tau_i \frac{4\mu}{9L} + 1 - \tau_i)^{1/k}} \left(e^{\frac{13}{6} \frac{n}{k} \ln \frac{2L}{\mu}} - 1\right)\right]^{k/2} \sqrt{\frac{3L}{2\mu}} \cdot \lambda_0. \tag{50}$$

---

[3] Hereinafter, $\lceil t \rceil$ for $t > 0$ denotes the smallest positive integer greater or equal to t.

Proof Let us prove by induction that, for all $k \geq 0$, we have

$$\xi_k \leq \tfrac{3}{2}. \tag{51}$$

Clearly, (51) is satisfied for $k = 0$ since $\xi_0 = 1$. It is also satisfied for $k = 1$ since $\xi_1 \overset{(31)}{=} e^{Mr_0} \overset{(35)}{\leq} e^{\xi_0 M\lambda_0} \overset{(31)}{=} e^{M\lambda_0} \overset{(47)}{\leq} \tfrac{3}{2}$.

Now let $k \geq 0$, and suppose that (51) has already been proved for all indices up to $k + 1$. Then, applying Lemma 5.2, we obtain (48) for all indices up to $k + 1$. Applying now Lemma 5.3 and using for all $0 \leq i \leq k$ the relation $q_i \overset{(37)}{=} \max\{1 - \frac{\mu}{\xi_{i+1}L}, \xi_{i+1} - 1\} \overset{(51)}{\leq} \max\{1 - \frac{2\mu}{3L}, \frac{1}{2}\} \leq 1 - \frac{\mu}{2L}$, we obtain (49) for all indices up to $k + 1$. Finally, if $k \geq 1$, then, applying Lemma 5.4 and using that $\xi_{i+1}^{\xi_{i+1}} \overset{(51)}{\leq} (\tfrac{3}{2})^{\frac{3}{2}} = \tfrac{3}{2}\sqrt{\tfrac{3}{2}} \leq \tfrac{3}{2}(1 + \tfrac{1}{4}) = \tfrac{15}{8} \leq 2$ for all $0 \leq i \leq k$, we obtain (50) for all indices up to k. Thus, at this moment, (48) and (49) are proved for all indices up to $k + 1$, while (50) is proved only up to k.

To finish the inductive step, it remains to prove that (51) is satisfied for the index $k + 2$, or, equivalently, in view of (31), that $M \sum_{i=0}^{k+1} r_i \leq \ln\tfrac{3}{2}$. Since $M \sum_{i=0}^{k+1} r_i \leq M \sum_{i=0}^{k+1} \xi_i \lambda_i \leq \tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i$ in view of (35) and (51) respectively, it suffices to show that $\tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i \leq \ln\tfrac{3}{2}$.

Note that

$$\tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i \overset{(49)}{\leq} \left(\tfrac{3}{2}\right)^{\frac{3}{2}} M\lambda_0 \sum_{i=0}^{k+1} \left(1 - \tfrac{\mu}{2L}\right)^i \leq \left(\tfrac{3}{2}\right)^{\frac{3}{2}} \tfrac{2L}{\mu} M\lambda_0. \tag{52}$$

Therefore, if we could prove that

$$\tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i \leq \left(\tfrac{3}{2}\right)^{\frac{3}{2}} (K_0 + 9) M\lambda_0, \tag{53}$$

then, combining (52) and (53), we would obtain

$$\tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i \leq \left(\tfrac{3}{2}\right)^{\frac{3}{2}} \min\left\{\tfrac{2L}{\mu}, K_0 + 9\right\} M\lambda_0 \overset{(47)}{\leq} \ln\tfrac{3}{2},$$

which is exactly what we need. Let us prove (53). If $k \leq K_0$, in view of (49), we have $\tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i \leq \left(\tfrac{3}{2}\right)^{\frac{3}{2}} (k+2) M\lambda_0 \leq \left(\tfrac{3}{2}\right)^{\frac{3}{2}} (K_0 + 2) M\lambda_0$, and (53) follows. Therefore, from now on, we can assume that $k \geq K_0$. Then[4],

$$
\begin{aligned}
\tfrac{3}{2} M \sum_{i=0}^{k+1} \lambda_i &= \tfrac{3}{2} M \left(\sum_{i=0}^{K_0-1} \lambda_i + \lambda_{k+1}\right) + \tfrac{3}{2} M \sum_{i=K_0}^{k} \lambda_i \\
&\overset{(49)}{\leq} \left(\tfrac{3}{2}\right)^{\frac{3}{2}} (K_0 + 1) M\lambda_0 + \tfrac{3}{2} M \sum_{i=K_0}^{k} \lambda_i.
\end{aligned}
$$

It remains to show $\tfrac{3}{2} M \sum_{i=K_0}^{k} \lambda_i \leq \left(\tfrac{3}{2}\right)^{\frac{3}{2}} 8 M\lambda_0$. We can do this using (50).

---

[4] We will estimate the second sum using (50). However, recall that, at this moment, (50) is proved only up to the index k. This is the reason why we move λ_{k+1} into the first sum.

First, let us make some estimations. Clearly, for all $0 < t < 1$, we have $e^t = \sum_{j=0}^{\infty} \frac{t^j}{j!} \le 1 + t + \frac{t^2}{2} \sum_{j=0}^{\infty} t^j = 1 + t(1 + \frac{t}{2(1-t)})$. Hence, for all $0 < t \le 1$, we obtain $e^{\frac{13t}{48}} - 1 \le \frac{13t}{48}(1 + \frac{\frac{13}{48}}{2(1-\frac{13}{48})}) = \frac{13t}{48} \cdot \frac{83}{70} \le \frac{13t}{48} \cdot \frac{6}{5} = \frac{13t}{40}$, and so

$$\left[\frac{5}{2t} \left(e^{\frac{13t}{48}} - 1 \right) \right]^{1/2} \le \sqrt{\frac{5}{2t} \cdot \frac{13t}{40}} = \sqrt{\frac{13}{16}} \le \frac{11}{12}. \tag{54}$$

At the same time, $\frac{11}{12} = 1 - \frac{1}{12} \le e^{-\frac{1}{12}}$. Hence,

$$\left(\frac{11}{12}\right)^{K_0} \sqrt{\frac{L}{\mu}} \overset{(45)}{\le} \left(\frac{11}{12}\right)^{8 \ln \frac{2L}{\mu}} \sqrt{\frac{L}{\mu}} \le e^{-\frac{2}{3} \ln \frac{2L}{\mu}} \sqrt{\frac{L}{\mu}} = \left(\frac{2L}{\mu}\right)^{-\frac{2}{3}} \sqrt{\frac{L}{\mu}} \tag{55}$$
$$= 2^{-\frac{2}{3}} \left(\frac{L}{\mu}\right)^{-\frac{1}{6}} \le 2^{-\frac{2}{3}} \le \frac{2}{3}.$$

Thus, for all $K_0 \le i \le k$, and $p := \tau \frac{4\mu}{9L} + 1 - \tau \overset{(45)}{\le} \prod_{j=0}^{i-1}(\tau_i \frac{4\mu}{9L} + 1 - \tau_i)^{1/i}$:

$$\lambda_i \overset{(50)}{\le} \left[\frac{5}{2p} \left(e^{\frac{13}{6} \frac{n}{i} \ln \frac{2L}{\mu}} - 1 \right) \right]^{i/2} \sqrt{\frac{3L}{2\mu}} \cdot \lambda_0$$
$$\overset{(45)}{\le} \left[\frac{5}{2p} \left(e^{\frac{13p}{48}} - 1 \right) \right]^{i/2} \sqrt{\frac{3L}{2\mu}} \cdot \lambda_0 \overset{(54)}{\le} \left(\frac{11}{12}\right)^i \sqrt{\frac{3L}{2\mu}} \cdot \lambda_0$$
$$= \left(\frac{11}{12}\right)^{i-K_0} \left(\frac{11}{12}\right)^{K_0} \sqrt{\frac{3L}{2\mu}} \cdot \lambda_0 \overset{(55)}{\le} \left(\frac{11}{12}\right)^{i-K_0} \frac{2}{3} \cdot \sqrt{\frac{3}{2}} \cdot \lambda_0.$$

Hence, $\frac{3}{2} M \sum_{i=K_0}^{k} \lambda_i \le (\frac{3}{2})^{\frac{3}{2}} M \lambda_0 \cdot \frac{2}{3} \sum_{i=K_0}^{k} (\frac{11}{12})^{i-K_0} \le (\frac{3}{2})^{\frac{3}{2}} 8 M \lambda_0$. $\qquad \square$

Remark 5.2 In accordance with Theorem 5.1, the parameter M of strong self-concordancy affects only the size of the region of local convergence of the process (25), and not its rate of convergence. We do not know whether this is an artifact of the analysis or not, but it might be an interesting topic for future research. For a quadratic function, we have $M = 0$, and so the scheme (25) is globally convergent.

The region of local convergence, specified by (47), depends on the *maximum* of two quantities: $\frac{\mu}{L}$ and $\frac{1}{K_0}$. For DFP, the $\frac{1}{K_0}$ part in this maximum is in fact redundant, and its region of local convergence is simply inversely proportional to the condition number: $O\left(\frac{\mu}{L}\right)$. However, for BFGS, the $\frac{1}{K_0}$ part does not disappear, and we obtain the following region of local convergence:

$$M \lambda_0 \le \max \left\{ O\left(\frac{\mu}{L}\right), \, O\left(\frac{1}{n \ln \frac{2L}{\mu}}\right) \right\}.$$

Clearly, the latter region can be much bigger than the former when the condition number $\frac{L}{\mu}$ is significantly larger than the dimension n.

Remark 5.3 The previous estimate of the size of the region of local convergence, established in [27], was $O(\frac{\mu}{L})$ for both DFP and BFGS.

Example 5.1 Consider the functions

$$f(x) := f_0(x) + \tfrac{\mu}{2}\|x\|^2, \qquad f_0(x) := \ln\left(\sum_{i=1}^{m} e^{\langle a_i, x\rangle + b_i}\right), \qquad x \in \mathbb{E},$$

where $a_i \in \mathbb{E}^*$, $b_i \in \mathbb{R}$, $i = 1, \ldots, m$, $\mu > 0$, and $\|\cdot\|$ is the Euclidean norm, induced by the operator B. Let $\gamma > 0$ be such that

$$\|a_i\|_* \leq \gamma, \qquad i = 1, \ldots, m,$$

where $\|\cdot\|_*$ is the norm conjugate to $\|\cdot\|$. Define

$$\pi_i(x) := \frac{e^{\langle a_i, x\rangle + b_i}}{\sum_{j=1}^{m} e^{\langle a_j, x\rangle + b_j}}, \qquad x \in \mathbb{E}, \quad i = 1, \ldots, m.$$

Clearly, $\sum_{i=1}^{m} \pi_i(x) = 1$, $\pi_i(x) > 0$ for all $x \in \mathbb{E}$, $i = 1, \ldots, m$. It is not difficult to check that, for all $x, h \in \mathbb{E}$, we have[5]

$$
\begin{aligned}
\langle \nabla f_0(x), h\rangle &= \sum_{i=1}^{m} \pi_i(x)\langle a_i, h\rangle \;\leq\; \gamma. \\
\langle \nabla^2 f_0(x)h, h\rangle &= \sum_{i=1}^{m} \pi_i(x)\langle a_i - \nabla f_0(x), h\rangle^2 \\
&= \sum_{i=1}^{m} \pi_i(x)\langle a_i, h\rangle^2 - \langle \nabla f_0(x), h\rangle^2 \;\leq\; \gamma^2\|h\|^2, \\
D^3 f_0(x)[h, h, h] &= \sum_{i=1}^{m} \pi_i(x)\langle a_i - \nabla f_0(x), h\rangle^3 \\
&\leq 2\gamma\|h\|\langle \nabla^2 f_0(x)h, h\rangle \;\leq\; 2\gamma^3\|h\|^3.
\end{aligned}
$$

Thus, f_0 is a convex function with γ^2-Lipschitz gradient and $(2\gamma^3)$-Lipschitz Hessian. Consequently, the function f is μ-strongly convex with L-Lipschitz gradient, $(2\gamma^3)$-Lipschitz Hessian, and, in view of [26, Example 4.1], M-strongly self-concordant, where

$$L := \gamma^2 + \mu, \qquad M := \tfrac{2\gamma^3}{\mu^{3/2}}.$$

Let the regularization parameter μ be sufficiently small, namely $\mu \leq \gamma^2$. Denote $Q := \frac{\gamma^2}{\mu} \geq 1$. Then, $Q \leq \frac{L}{\mu} \leq 2Q$, $M = 2Q^{3/2}$, so, according to (47), the region of local convergence of BFGS can be described as follows:

$$\lambda_0 \leq \max\left\{ O\left(\tfrac{1}{Q^{5/2}}\right), O\left(\tfrac{1}{nQ^{3/2}\ln(4Q)}\right)\right\}. \qquad\qquad \square$$

---

[5] $D^3 f_0(x)[h, h, h] = \frac{d^3}{dt^3} f_0(x + th)\big|_{t=0}$ is the third derivative of f along the direction h.

6 Discussion

Let us compare the new convergence rates, obtained in this paper for the classical DFP and BFGS methods, with the previously known ones from [27]. Since the estimates for the general nonlinear case differ from those for the quadratic one just in absolute constants, we only discuss the latter case.

In what follows, we use our standard notation: n is the dimension of the space, μ is the strong convexity parameter, L is the Lipschitz constant of the gradient, and λ_k is the local norm of the gradient at the kth iteration.

For BFGS, the previously known rate (see [27, Theorem 3.2]) is

$$\lambda_k \leq \left(\tfrac{nL}{\mu k}\right)^{k/2} \lambda_0. \tag{56}$$

Although (56) is formally valid for all $k \geq 1$, it becomes useful[6] only after

$$\widehat{K}_0^{\mathrm{BFGS}} := \tfrac{nL}{\mu} \tag{57}$$

iterations. Thus, $\widehat{K}_0^{\mathrm{BFGS}}$ can be thought of as the *starting moment* of the superlinear convergence, according to the estimate (56).

In this paper, we have obtained a new estimate (Theorem 4.2):

$$\lambda_k \leq \left[2\left(e^{\frac{n}{k}\ln\frac{L}{\mu}} - 1\right)\right]^{k/2} \sqrt{\tfrac{L}{\mu}} \cdot \lambda_0. \tag{58}$$

Its starting moment of superlinear convergence can be described as follows:

$$K_0^{\mathrm{BFGS}} := 4n \ln \tfrac{L}{\mu}. \tag{59}$$

Indeed, since $e^t \leq \frac{1}{1-t} = 1 + \frac{t}{1-t}$ for any $t < 1$, we have, for all $k \geq K_0^{\mathrm{BFGS}}$,

$$e^{\frac{n}{k}\ln\frac{L}{\mu}} - 1 \leq \frac{\frac{n}{k}\ln\frac{L}{\mu}}{1-\frac{n}{k}\ln\frac{L}{\mu}} \overset{(59)}{\leq} \frac{\frac{n}{k}\ln\frac{L}{\mu}}{1-\frac{1}{4}} = \frac{4n}{3k}\ln\tfrac{L}{\mu}. \tag{60}$$

At the same time, for all $k \geq K_0^{\mathrm{BFGS}}$:

$$\sqrt{\tfrac{L}{\mu}} = e^{\frac{1}{2}\ln\frac{L}{\mu}} \overset{(59)}{\leq} e^{\frac{k}{8}} = (e^{\frac{1}{4}})^{k/2} \leq \left(\tfrac{4}{3}\right)^{k/2} \leq \left(\tfrac{3}{2}\right)^{k/2}. \tag{61}$$

Hence, according the new estimate (58), for all $k \geq K_0^{\mathrm{BFGS}}$:

$$\lambda_k \overset{(60)}{\leq} \left(\tfrac{8n}{3k}\ln\tfrac{L}{\mu}\right)^{k/2} \sqrt{\tfrac{L}{\mu}} \cdot \lambda_0 \overset{(61)}{\leq} \left(\tfrac{4n}{k}\ln\tfrac{L}{\mu}\right)^{k/2} \lambda_0 \qquad \left(\overset{(59)}{\leq} \lambda_0\right). \tag{62}$$

Comparing the previously known efficiency estimate (56) and its starting moment of superlinear convergence (57) with the new ones (62), (59), we thus conclude that we manage to put the condition number $\frac{L}{\mu}$ *under the logarithm*.

---

[6] Indeed, according to Theorem 4.1, we have at least $\lambda_k \leq (1 - \frac{\mu}{L})^k \lambda_0$ for all $k \geq 0$.

For DFP, the previously known rate (see [27, Theorem 3.2]) is

$$\lambda_k \leq \left(\frac{nL^2}{\mu^2 k}\right)^{k/2} \lambda_0$$

with the following starting moment of the superlinear convergence:

$$\widehat{K}_0^{\text{DFP}} := \frac{nL^2}{\mu^2}. \tag{63}$$

The new rate, which we have obtained in this paper (Theorem 4.2), is

$$\lambda_k \leq \left[\frac{2L}{\mu}\left(e^{\frac{n}{k}\ln\frac{L}{\mu}} - 1\right)\right]^{k/2} \sqrt{\frac{L}{\mu}} \cdot \lambda_0. \tag{64}$$

Repeating the same reasoning as above, we can easily obtain that the new starting moment of the superlinear convergence can be described as follows:

$$K_0^{\text{DFP}} := \frac{4nL}{\mu}\ln\frac{L}{\mu}, \tag{65}$$

and, for all $k \geq K_0^{\text{DFP}}$, the new estimate (64) takes the following form:

$$\lambda_k \leq \left(\frac{4nL}{\mu k}\ln\frac{L}{\mu}\right)^{k/2} \lambda_0 \quad (\overset{(65)}{\leq} \lambda_0).$$

Thus, compared to the old result, we have improved the factor $\frac{L^2}{\mu^2}$ up to $\frac{L}{\mu}\ln\frac{L}{\mu}$. Interestingly enough, the ratio between the old starting moments (63), (57) of the superlinear convergence of DFP and BFGS and the new ones (65), (59) have remained the same, $\frac{L}{\mu}$, although the both estimates have been improved.

It is also interesting whether the results, obtained in this paper, can be applied to *limited-memory* quasi-Newton methods such as L-BFGS [30]. Unfortunately, it seems like the answer is negative. The main problem is that we cannot say anything interesting about just a *few* iterations of BFGS. Indeed, according to our main result, after k iterations of BFGS, the initial residual is contracted by the factor of the form $[\exp(\frac{n}{k}\ln\frac{L}{\mu}) - 1]^k$. For all values $k \leq n\ln\frac{L}{\mu}$, this contraction factor is in fact bigger than 1, so the result becomes useless.

7 Conclusions

We have presented a new theoretical analysis of local superlinear convergence of classical quasi-Newton methods from the convex Broyden class. Our analysis has been based on the potential function involving the logarithm of determinant of Hessian approximation and the trace of inverse Hessian approximation. Compared to the previous works, we have obtained new convergence rate estimates, which have much better dependency on the condition number of the problem.

Note that all our results are *local*, i.e. they are valid under the assumption that the starting point is sufficiently close to a minimizer. In particular, there

is no contradiction between our results and the fact that the DFP method is not known to be globally convergent with inexact line search (see, e.g., [31]).

Let us mention several open questions. First, looking at the starting moment of superlinear convergence of the BFGS method, in addition to the dimension of the problem, we see the presence of the logarithm of its condition number. Although typically such logarithmic factors are considered small, it is still interesting to understand whether this factor can be completely removed.

Second, all the superlinear convergence rates, which we have obtained for the convex Broyden class in this paper, are expressed in terms of the parameter τ, which controls the weight of the DFP component in the updating formula for the *inverse* operator. At the same time, in [27], the corresponding estimates were presented in terms of the parameter ϕ, which controls the weight of the DFP component in the updating formula for the *primal* operator. Of course, for the extreme members of the convex Broyden class, DFP and BFGS, ϕ and τ coincide. However, in general, they could be quite different. We do not know if it is possible to express the results of this paper in terms of ϕ instead of τ.

Finally, in all the methods, which we considered, the initial Hessian approximation G_0 was LB, where L is the Lipschitz constant of the gradient, measured relative to the operator B. We always assume that this constant is known. Of course, it is interesting to develop some *adaptive* algorithms, which could start from any initial guess L_0 for the constant L, and then somehow dynamically adjust the Hessian approximations in iterations, yet retaining all the original efficiency estimates.

Appendix

Lemma A.1 *Let $A, G : \mathbb{E} \to \mathbb{E}^*$ be self-adjoint positive definite linear operators, let $u \in \mathbb{E}$ be non-zero, and let $\tau \in \mathbb{R}$ be such that $G_+ := \mathrm{Broyd}_\tau(A, G, u)$ is well-defined. Then,*

$$G_+^{-1} = \tau \left[G^{-1} - \frac{G^{-1}Auu^*AG^{-1}}{\langle AG^{-1}Au, u \rangle} + \frac{uu^*}{\langle Au, u \rangle} \right]$$
$$+ (1 - \tau) \left[G^{-1} - \frac{G^{-1}Auu^* + uu^*AG^{-1}}{\langle Au, u \rangle} + \left(\frac{\langle AG^{-1}Au, u \rangle}{\langle Au, u \rangle} + 1 \right) \frac{uu^*}{\langle Au, u \rangle} \right], \tag{66}$$

and

$$\mathrm{Det}(G_+^{-1}, G) = \tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Gu, u \rangle}{\langle Au, u \rangle}. \tag{67}$$

Proof Denote $\phi := \phi_\tau(A, G, u)$. According to Lemma 6.2 in [27], we have

$$\mathrm{Det}(G^{-1}, G_+) = \phi \frac{\langle AG^{-1}Au, u \rangle}{\langle Au, u \rangle} + (1 - \phi) \frac{\langle Au, u \rangle}{\langle Gu, u \rangle} \stackrel{(3)}{=} \left[\tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} \right]^{-1}.$$

This proves (67) since $\mathrm{Det}(G_+^{-1}, G) = \frac{1}{\mathrm{Det}(G^{-1}, G_+)}$. Let us prove (66). Denote

$$G_0 := G - \frac{Guu^*G}{\langle Gu, u \rangle} + \frac{Auu^*A}{\langle Au, u \rangle}, \qquad s := \frac{Au}{\langle Au, u \rangle} - \frac{Gu}{\langle Gu, u \rangle}. \tag{68}$$

Note that

$$G_+ \stackrel{(2)}{=} G_0 + \phi \left[\frac{\langle Gu, u \rangle Auu^*A}{\langle Au, u \rangle^2} + \frac{Guu^*G}{\langle Gu, u \rangle} - \frac{Auu^*G + Guu^*A}{\langle Au, u \rangle} \right] = G_0 + \phi \langle Gu, u \rangle ss^*. \tag{69}$$

Let $I_{\mathbb{E}}$ and $I_{\mathbb{E}^*}$ be the identity operators in $\mathbb{E}$ and $\mathbb{E}^*$. Since $G_0 u = Au$, we have

$$
\begin{aligned}
&\left[\left(I_{\mathbb{E}} - \frac{uu^*A}{\langle Au, u \rangle} \right) G^{-1} \left(I_{\mathbb{E}^*} - \frac{Auu^*}{\langle Au, u \rangle} \right) + \frac{uu^*}{\langle Au, u \rangle} \right] G_0 \\
&= \left(I_{\mathbb{E}} - \frac{uu^*A}{\langle Au, u \rangle} \right) G^{-1} \left(G_0 - \frac{Auu^*A}{\langle Au, u \rangle} \right) + \frac{uu^*A}{\langle Au, u \rangle} \\
&\overset{(68)}{=} \left(I_{\mathbb{E}} - \frac{uu^*A}{\langle Au, u \rangle} \right) G^{-1} \left(G - \frac{Guu^*G}{\langle Gu, u \rangle} \right) + \frac{uu^*A}{\langle Au, u \rangle} = I_{\mathbb{E}}.
\end{aligned}
$$

Hence, we can conclude that

$$
\begin{aligned}
G_0^{-1} &= \left(I_{\mathbb{E}} - \frac{uu^*A}{\langle Au, u \rangle} \right) G^{-1} \left(I_{\mathbb{E}^*} - \frac{Auu^*}{\langle Au, u \rangle} \right) + \frac{uu^*}{\langle Au, u \rangle} \\
&= G^{-1} - \frac{G^{-1}Auu^* + uu^*AG^{-1}}{\langle Au, u \rangle} + \left(\frac{\langle AG^{-1}Au, u \rangle}{\langle Au, u \rangle} + 1 \right) \frac{uu^*}{\langle Au, u \rangle}.
\end{aligned}
$$

Thus, we see that the right-hand side of (66) equals

$$
\begin{aligned}
H_+ &:= G_0^{-1} - \tau \left[\frac{\langle AG^{-1}Au, u \rangle uu^*}{\langle Au, u \rangle^2} + \frac{G^{-1}Auu^*AG^{-1}}{\langle AG^{-1}Au, u \rangle} - \frac{G^{-1}Auu^* + uu^*AG^{-1}}{\langle Au, u \rangle} \right] \\
&= G_0^{-1} - \tau \langle AG^{-1}Au, u \rangle ww^*,
\end{aligned}
\tag{70}
$$

where

$$
w := \frac{G^{-1}Au}{\langle AG^{-1}Au, u \rangle} - \frac{u}{\langle Au, u \rangle}.
\tag{71}
$$

It remains to verify that $H_+ G_+ = I_{\mathbb{E}}$. Clearly,

$$
\begin{aligned}
\langle AG^{-1}Au, u \rangle G_0 w &\overset{(71)}{=} G_0 G^{-1}Au - \frac{\langle AG^{-1}Au, u \rangle G_0 u}{\langle Au, u \rangle} \\
&\overset{(68)}{=} Au - \frac{\langle Au, u \rangle Gu}{\langle Gu, u \rangle} \overset{(68)}{=} \langle Au, u \rangle s.
\end{aligned}
\tag{72}
$$

Hence,

$$
\begin{aligned}
\langle AG^{-1}Au, u \rangle \langle G_0 w, w \rangle &\overset{(72)}{=} \langle Au, u \rangle \langle s, w \rangle \overset{(71)}{=} \frac{\langle Au, u \rangle \langle s, G^{-1}Au \rangle}{\langle AG^{-1}Au, u \rangle} - \langle s, u \rangle \\
&\overset{(68)}{=} \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} \left(\frac{\langle AG^{-1}Au, u \rangle}{\langle Au, u \rangle} - \frac{\langle Au, u \rangle}{\langle Gu, u \rangle} \right) = 1 - \frac{\langle Au, u \rangle^2}{\langle AG^{-1}Au, u \rangle \langle Gu, u \rangle}.
\end{aligned}
\tag{73}
$$

Consequently,

$$
\begin{aligned}
\frac{\langle Gu, u \rangle}{\langle Au, u \rangle} H_+ G_0 ww^* G_0 &\overset{(70)}{=} \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} (G_0^{-1} - \tau \langle AG^{-1}Au, u \rangle ww^*) G_0 ww^* G_0 \\
&= \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} (1 - \tau \langle AG^{-1}Au, u \rangle \langle G_0 w, w \rangle) ww^* G_0 \\
&\overset{(73)}{=} \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} \left(1 - \tau + \tau \frac{\langle Au, u \rangle^2}{\langle AG^{-1}Au, u \rangle \langle Gu, u \rangle} \right) ww^* G_0 \\
&= \left[\tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} \right] ww^* G_0.
\end{aligned}
\tag{74}
$$

Thus,

$$
\begin{aligned}
H_+ G_+ &\overset{(69)}{=} H_+ (G_0 + \phi \langle Gu, u \rangle ss^*) \overset{(72)}{=} H_+ \left(G_0 + \phi \frac{\langle AG^{-1}Au, u \rangle^2}{\langle Au, u \rangle} \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} G_0 ww^* G_0 \right) \\
&\overset{(74)}{=} H_+ G_0 + \phi \frac{\langle AG^{-1}Au, u \rangle^2}{\langle Au, u \rangle} \left[\tau \frac{\langle Au, u \rangle}{\langle AG^{-1}Au, u \rangle} + (1 - \tau) \frac{\langle Gu, u \rangle}{\langle Au, u \rangle} \right] \\
&\overset{(3)}{=} H_+ G_0 + \tau \langle AG^{-1}Au, u \rangle ww^* G_0 \overset{(70)}{=} I_{\mathbb{E}}. \qquad \square
\end{aligned}
$$

References

1. Davidon, W.: Variable metric method for minimization. Argonne National Laboratory Research and Development Report 5990 (1959)
2. Fletcher, R., Powell, M.: A rapidly convergent descent method for minimization. Computer Journal. **6**(2), 163–168 (1963)
3. Broyden, C.: The convergence of a class of double-rank minimization algorithms: 1. General considerations. IMA Journal of Applied Mathematics. **6**(1), 76–90 (1970)
4. Broyden, C.: The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. IMA Journal of Applied Mathematics. **6**(3), 222–231 (1970)
5. Fletcher, R.: A new approach to variable metric algorithms. Computer Journal. **13**(3), 317–322 (1970)
6. Goldfarb, D.: A family of variable-metric methods derived by variational means. Mathematics of Computation. **24**(109), 23–26 (1970)
7. Shanno, D.: Conditioning of quasi-Newton methods for function minimization. Mathematics of Computation. **24**(111), 647–656 (1970)
8. Broyden, C.: Quasi-Newton methods and their application to function minimization. Mathematics of Computation. **21**(99), 368–381 (1967)
9. Dennis, J., Moré, J.: Quasi-Newton methods, motivation and theory. SIAM Review. **19**(1), 46–89 (1977)
10. Nocedal, J., Wright, S.: Numerical optimization. Springer Science & Business Media, New York, NY, USA (2006)
11. Lewis, A., Overton, M.: Nonsmooth optimization via quasi-Newton methods. Mathematical Programming. **141**(1-2), 135–163 (2013)
12. Powell, M.: On the convergence of the variable metric algorithm. IMA Journal of Applied Mathematics. **7**(1), 21–36 (1971)
13. Dixon, L.: Quasi-Newton algorithms generate identical points. Mathematical Programming. **2**(1), 383–387 (1972)
14. Dixon, L.: Quasi Newton techniques generate identical points II: The proofs of four new theorems. Mathematical Programming. **3**(1), 345–358 (1972)
15. Broyden, C., Dennis, J., Moré, J.: On the local and superlinear convergence of quasi-Newton methods. IMA Journal of Applied Mathematics. **12**(3), 223–245 (1973)
16. Dennis, J., Moré, J.: A characterization of superlinear convergence and its application to quasi-Newton methods. Mathematics of Computation. **28**(126), 549–560 (1974)
17. Stachurski, A.: Superlinear convergence of Broyden's bounded θ-class of methods. Mathematical Programming. **20**(1), 196–212 (1981)
18. Griewank, A., Toint, P.: Local convergence analysis for partitioned quasi-Newton updates. Numerische Mathematik. **39**(3), 429–448 (1982)
19. Engels, J., Martínez, H.: Local and superlinear convergence for partially known quasi-Newton methods. SIAM Journal on Optimization. **1**(1), 42–56 (1991)
20. Byrd, R., Liu, D., Nocedal, J.: On the behavior of Broyden's class of quasi-Newton methods. SIAM Journal on Optimization. **2**(4), 533–557 (1992)
21. Yabe, H., Yamaki, N.: Local and superlinear convergence of structured quasi-Newton methods for nonlinear optimization. Journal of the Operations Research Society of Japan. **39**(4), 541–557 (1996)
22. Wei, Z., Yu, G., Yuan, G., Lian, Z.: The superlinear convergence of a modified BFGS-type method for unconstrained optimization. Computational Optimization and Applications. **29**(3), 315–332 (2004)
23. Yabe, H., Ogasawara, H., Yoshino, M.: Local and superlinear convergence of quasi-Newton methods based on modified secant conditions. Journal of Computational and Applied Mathematics. **205**(1), 617–632 (2007)
24. Mokhtari, A., Eisen, M., Ribeiro, A.: IQN: An incremental quasi-Newton method with local superlinear convergence rate. SIAM Journal on Optimization. **28**(2), 1670–1698 (2018)
25. Gao, W., Goldfarb, D.: Quasi-Newton methods: superlinear convergence without line searches for self-concordant functions. Optimization Methods and Software. **34**(1), 194–217 (2019)

26. Rodomanov, A., Nesterov, Y.: Greedy quasi-Newton methods with explicit superlinear convergence. CORE Discussion Papers. **06** (2020)
27. Rodomanov, A., Nesterov, Y.: Rates of Superlinear Convergence for Classical Quasi-Newton Methods. CORE Discussion Papers. **11** (2020)
28. Jin, Q., Mokhtari, A.: Non-asymptotic Superlinear Convergence of Standard Quasi-Newton Methods. arXiv preprint arXiv:2003.13607 (2020)
29. Byrd, R., Nocedal, J.: A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. SIAM Journal on Numerical Analysis. **26**(3), 727–739 (1989)
30. Liu, D., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Mathematical Programming. **45**(1-3), 503–528 (1989).
31. Byrd, R., Nocedal, J., Yuan, Y.: Global convergence of a class of quasi-Newton methods on convex problems. SIAM Journal on Numerical Analysis. **24**(5), 1171–1190 (1987).