Rates of Convergence

Newton’s Method
Rates of Convergence

We compare the performance of algorithms by their *rate of convergence*. 
Rates of Convergence

We compare the performance of algorithms by their rate of convergence.

That is, if $x^k \to \bar{x}$, we are interested in how fast this happens.
We compare the performance of algorithms by their rate of convergence.

That is, if $x^k \to \bar{x}$, we are interested in how fast this happens.

We consider only quotient rates, or Q-rates of convergence.
Let \( \{x^\nu\} \subset \mathbb{R}^n \) and \( \bar{x} \in \mathbb{R}^n \) be such that \( \bar{x}^\nu \to \bar{x} \).
Let \( \{x^\nu\} \subset \mathbb{R}^n \) and \( \bar{x} \in \mathbb{R}^n \) be such that \( x^\nu \to \bar{x} \).

We say that \( x^\nu \to \bar{x} \) at a *linear* rate if

\[
\limsup_{\nu \to \infty} \frac{\|x^{\nu+1} - \bar{x}\|}{\|x^\nu - \bar{x}\|} < 1.
\]
Let \( \{x^\nu\} \subset \mathbb{R}^n \) and \( \bar{x} \in \mathbb{R}^n \) be such that \( \bar{x}^\nu \to \bar{x} \).

We say that \( \bar{x}^\nu \to \bar{x} \) at a **linear** rate if

\[
\limsup_{\nu \to \infty} \frac{\|x^{\nu+1} - \bar{x}\|}{\|x^\nu - \bar{x}\|} < 1.
\]

The convergence is said to be **superlinear** if this limsup is 0.
Let \( \{x^\nu\} \subset \mathbb{R}^n \) and \( \bar{x} \in \mathbb{R}^n \) be such that \( x^\nu \to \bar{x} \).

We say that \( x^\nu \to \bar{x} \) at a *linear* rate if

\[
\limsup_{\nu \to \infty} \frac{\|x^{\nu+1} - \bar{x}\|}{\|x^\nu - \bar{x}\|} < 1.
\]

The convergence is said to be *superlinear* if this limsup is 0.

The convergence is said to be *quadratic* if

\[
\limsup_{\nu \to \infty} \frac{\|x^{\nu+1} - \bar{x}\|}{\|x^\nu - \bar{x}\|^2} < \infty.
\]
Rates of Convergence: Example

Let $\gamma \in (0, 1)$. 
\{\gamma^n\} converges linearly to zero, but not superlinearly.
Rates of Convergence: Example

Let $\gamma \in (0, 1)$.

$\{\gamma^n\}$ converges linearly to zero, but not superlinearly.

$\{\gamma^{n^2}\}$ converges superlinearly to 0, but not quadratically.
Let $\gamma \in (0, 1)$.

- $\{\gamma^n\}$ converges linearly to zero, but not superlinearly.
- $\{\gamma^{n^2}\}$ converges superlinearly to 0, but not quadratically.
- $\{\gamma^{2^n}\}$ converges quadratically to zero.
Rates of Convergence: Example

Let $\gamma \in (0, 1)$.

$\{\gamma^n\}$ converges linearly to zero, but not superlinearly.

$\{\gamma^{n^2}\}$ converges superlinearly to 0, but not quadratically.

$\{\gamma^{2^n}\}$ converges quadratically to zero.

Superlinear convergence is much faster than linear convergences, but quadratic convergence is much, much faster than superlinear convergence.
Let $\gamma \in (0, 1)$. 

$\{\gamma^n\}$ converges linearly to zero, but not superlinearly.

$\{\gamma^{n^2}\}$ converges superlinearly to 0, but not quadratically.

$\{\gamma^{2^n}\}$ converges quadratically to zero.

Superlinear convergence is much faster than linear convergences, but quadratic convergence is much, much faster than superlinear convergence.

$$
\gamma = \frac{1}{2} \text{ gives } \gamma^n = 2^{-n}, \quad \gamma^{n^2} = 2^{-n^2}, \quad \gamma^{2^n} = 2^{-2^n}
$$
Example

Let \( f(x) = x^2 + e^x \).

\( f \) is a strongly convex function with

\[
\begin{align*}
    f(x) &= x^2 + e^x \\
    f'(x) &= 2x + e^x \\
    f''(x) &= 2 + e^x > 2 \\
    f'''(x) &= e^x.
\end{align*}
\]

If we apply the steepest descent algorithm with backtracking (\( \gamma = 1/2, \ c = 0.01 \)) initiated at \( x^0 = 1 \).
### Example: Steepest Descent

<table>
<thead>
<tr>
<th>$k$</th>
<th>$x^k$</th>
<th>$f(x^k)$</th>
<th>$f'(x^k)$</th>
<th>$s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.37182818</td>
<td>4.7182818</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>0.8565307</td>
<td>-0.3934693</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>-0.25</td>
<td>0.8413008</td>
<td>0.2788008</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>-0.375</td>
<td>0.8279143</td>
<td>-0.0627107</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>-0.34075</td>
<td>0.8273473</td>
<td>0.0297367</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>-0.356375</td>
<td>0.8272131</td>
<td>-0.01254</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>-0.3485625</td>
<td>0.8271976</td>
<td>0.0085768</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>-0.3524688</td>
<td>0.8271848</td>
<td>-0.001987</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>-0.3514922</td>
<td>0.8271841</td>
<td>0.0006528</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>-0.3517364</td>
<td>0.827184</td>
<td>-0.000072</td>
<td>12</td>
</tr>
</tbody>
</table>
Example: Newton’s Method

\[ \min f(x) := x^2 + e^x \]

\[ x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} \]
Example: Newton’s Method

\[ \min f(x) := x^2 + e^x \]

\[ x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} \]

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f'(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.7182818</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$-1/3$</td>
<td>0.0498646</td>
</tr>
<tr>
<td>$-0.3516893$</td>
<td>0.00012</td>
</tr>
<tr>
<td>$-0.3517337$</td>
<td>0.000000000064</td>
</tr>
</tbody>
</table>
Example: Newton’s Method

\[ \min f(x) := x^2 + e^x \]

\[ x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} \]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f'(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.7182818</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-1/3</td>
<td>.0498646</td>
</tr>
<tr>
<td>-.3516893</td>
<td>.00012</td>
</tr>
<tr>
<td>-.3517337</td>
<td>.000000000064</td>
</tr>
</tbody>
</table>

In addition, one more iteration gives \( |f'(x^5)| \leq 10^{-20} \).
Newton’s Method: the Gold Standard

Newton’s method is an algorithm for solving nonlinear equations.
Newton’s Method: the Gold Standard

Newton’s method is an algorithm for solving nonlinear equations.

Given $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, find $x \in \mathbb{R}^n$ for which $g(x) = 0$. 
Newton’s Method: the Gold Standard

Newton’s method is an algorithm for solving nonlinear equations.

Given $g : \mathbb{R}^n \to \mathbb{R}^n$, find $x \in \mathbb{R}^n$ for which $g(x) = 0$.

Linearize and Solve:
Newton’s Method: the Gold Standard

Newton’s method is an algorithm for solving nonlinear equations.

Given $g : \mathbb{R}^n \to \mathbb{R}^n$, find $x \in \mathbb{R}^n$ for which $g(x) = 0$.

Linearize and Solve:
Given a current estimate of a solution $x^0$ obtain a new estimate $x^1$ as the solution to the equation

$$0 = g(x^0) + g'(x^0)(x - x^0),$$

and repeat.
Newton Like Methods

\[ x^{k+1} := x^k - [g'(x^k)]^{-1} g(x^k) \]
Newton Like Methods

\[ x^{k+1} := x^k - [g'(x^k)]^{-1} g(x^k) \]

Newton-Like Methods:

\[ x^{k+1} := x^k - J_k g(x^k) \]

where

\[ J_k \approx g'(x^k) \]
Convergence of Newton’s Method

Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable, $x^0 \in \mathbb{R}^n$, and $J_0 \in \mathbb{R}^{n \times n}$. Suppose that there exists $\bar{x}$, $x_0 \in \mathbb{R}^n$, and $\epsilon > 0$ with $\|x_0 - \bar{x}\| < \epsilon$ such that
Convergence of Newton’s Method

Let $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be differentiable, $x^0 \in \mathbb{R}^n$, and $J_0 \in \mathbb{R}^{n \times n}$. Suppose that there exists $\bar{x}$, $x_0 \in \mathbb{R}^n$, and $\epsilon > 0$ with $\|x_0 - \bar{x}\| < \epsilon$ such that

1. $g(\bar{x}) = 0,$
Convergence of Newton’s Method

Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable, $x^0 \in \mathbb{R}^n$, and $J_0 \in \mathbb{R}^{n \times n}$. Suppose that there exists $\bar{x}$, $x_0 \in \mathbb{R}^n$, and $\epsilon > 0$ with $\|x_0 - \bar{x}\| < \epsilon$ such that

1. $g(\bar{x}) = 0$,

2. $g'(x)^{-1}$ exists for $x \in B(\bar{x}; \epsilon) := \{x \in \mathbb{R}^n : \|x - \bar{x}\| < \epsilon\}$ with

$$\sup\{\|g'(x)^{-1}\| : x \in B(\bar{x}; \epsilon)\} \leq M_1$$
Convergence of Newton’s Method

Let \( g : \mathbb{R}^n \to \mathbb{R}^n \) be differentiable, \( x^0 \in \mathbb{R}^n \), and \( J_0 \in \mathbb{R}^{n \times n} \). Suppose that there exists \( \bar{x}, x_0 \in \mathbb{R}^n \), and \( \epsilon > 0 \) with \( \|x_0 - \bar{x}\| < \epsilon \) such that

1. \( g(\bar{x}) = 0 \),

2. \( g'(x)^{-1} \) exists for \( x \in B(\bar{x}; \epsilon) := \{x \in \mathbb{R}^n : \|x - \bar{x}\| < \epsilon\} \) with

   \[
   \sup\{\|g'(x)^{-1}\| : x \in B(\bar{x}; \epsilon)\} \leq M_1
   \]

3. \( g' \) is Lipschitz continuous on \( \text{cl}B(\bar{x}; \epsilon) \) with Lipschitz constant \( L \), and
Convergence of Newton’s Method

Let \( g : \mathbb{R}^n \rightarrow \mathbb{R}^n \) be differentiable, \( x^0 \in \mathbb{R}^n \), and \( J_0 \in \mathbb{R}^{n \times n} \). Suppose that there exists \( \bar{x}, x_0 \in \mathbb{R}^n \), and \( \epsilon > 0 \) with \( \|x_0 - \bar{x}\| < \epsilon \) such that

1. \( g(\bar{x}) = 0 \),

2. \( g'(x)^{-1} \) exists for \( x \in B(\bar{x}; \epsilon) := \{ x \in \mathbb{R}^n : \|x - \bar{x}\| < \epsilon \} \) with

\[
\sup\{\|g'(x)^{-1}\| : x \in B(\bar{x}; \epsilon)\} \leq M_1
\]

3. \( g' \) is Lipschitz continuous on \( \text{cl}B(\bar{x}; \epsilon) \) with Lipschitz constant \( L \), and

4. \( \theta_0 := \frac{LM_1}{2} \|x^0 - \bar{x}\| + M_0 K < 1 \) where \( K \geq \|(g'(x^0)^{-1} - J_0)y^0\| \), \( y^0 := g(x^0)/\|g(x^0)\| \), and \( M_0 = \max\{\|g'(x)\| : x \in B(\bar{x}; \epsilon)\} \).
Convergence of Newton’s Method

Further suppose that iteration is initiated at $x^0$ where the $J_k$’s are chosen to satisfy one of the following conditions:

1. $\|g'(x_k) - 1 - J_k y_k\| \leq K$,
2. $\|g'(x_k) - 1 - J_k y_k\| \leq \theta_k^1 K$ for some $\theta_1^k \in (0, 1)$,
3. $\|g'(x_k) - 1 - J_k y_k\| \leq \min\{M_2^2 \|x_k - x_{k-1}\|, K\}$, for some $M_2^2 > 0$,
4. $\|g'(x_k) - 1 - J_k y_k\| \leq \min\{M_3^2 \|g(x_k)\|, K\}$, for some $M_3^3 > 0$,

where for each $k = 1, 2, \ldots$, $y^k := g(x^k)/\|g(x^k)\|$.
Convergence of Newton’s Method

Further suppose that iteration is initiated at $x^0$ where the $J_k$’s are chosen to satisfy one of the following conditions:

(i) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq K$,

where for each $k = 1, 2, \ldots$, $y^k := g(x^k)/\|g(x^k)\|$. 
Further suppose that iteration is initiated at $x^0$ where the $J_k$'s are chosen to satisfy one of the following conditions:

(i) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq K$,

(ii) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq \theta_1 K$ for some $\theta_1 \in (0,1)$,

where for each $k = 1, 2, \ldots$, $y^k := g(x^k)/\|g(x^k)\|$.
Convergence of Newton’s Method

Further suppose that iteration is initiated at $x^0$ where the $J_k$’s are chosen to satisfy one of the following conditions;

(i) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq K$,

(ii) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq \theta_1^k K$ for some $\theta_1 \in (0, 1)$,

(iii) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq \min\{M_2 \| x^k - x^{k-1} \|, K \}$, for some $M_2 > 0$, or

where for each $k = 1, 2, \ldots$, $y^k := g(x^k)/\|g(x^k)\|$.
Convergence of Newton’s Method

Further suppose that iteration is initiated at $x^0$ where the $J_k$’s are chosen to satisfy one of the following conditions:

(i) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq K$,

(ii) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq \theta_1^k K$ for some $\theta_1 \in (0, 1)$,

(iii) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq \min\{M_2\|x^k - x^{k-1}\|, K\}$, for some $M_2 > 0$, or

(iv) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq \min\{M_2\|g(x^k)\|, K\}$, for some $M_3 > 0$,

where for each $k = 1, 2, \ldots$, $y^k := g(x^k)/\|g(x^k)\|$.
Convergence of Newton’s Method

These hypotheses on the accuracy of the approximations $J_k$ yield the following conclusions about the rate of convergence of the iterates $x^k$. 
Convergence of Newton’s Method

These hypotheses on the accuracy of the approximations $J_k$ yield the following conclusions about the rate of convergence of the iterates $x^k$.

(a) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq K \implies x^k \to \bar{x}$ linearly.
Convergence of Newton’s Method

These hypotheses on the accuracy of the approximations $J_k$ yield the following conclusions about the rate of convergence of the iterates $x^k$.

(a) $\| (g'(x^k)^{-1} - J_k)y^k \| \leq K \implies x^k \rightarrow \bar{x}$ linearly.

(b) $\| (g'(x^k)^{-1} - J_k)y^k \| \leq \theta_1^k K \implies x^k \rightarrow \bar{x}$ superlinearly.
Convergence of Newton’s Method

These hypotheses on the accuracy of the approximations $J_k$ yield the following conclusions about the rate of convergence of the iterates $x^k$.

(a) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq K \implies x^k \to \bar{x}$ linearly.

(b) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq \theta_1^k K \implies x^k \to \bar{x}$ superlinearly.

(c) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq \min\{M_2\|x^k - x^{k-1}\|, K\} \implies x^k \to \bar{x}$ two step quadratically.
Convergence of Newton’s Method

These hypotheses on the accuracy of the approximations $J_k$ yield the following conclusions about the rate of convergence of the iterates $x^k$.

(a) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq K \implies x^k \to \bar{x}$ linearly.

(b) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq \theta_1^k K \implies x^k \to \bar{x}$ superlinearly.

(c) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq \min\{M_2\|x^k - x^{k-1}\|, K\} \implies x^k \to \bar{x}$ two step quadratically.

(d) $\| (g'(x^k)^{-1} - J_k) y^k \| \leq \min\{M_2\|g(x^k)\|, K\} \implies x^k \to \bar{x}$ quadratically.
Newton’s Method for Minimization: $\nabla f(x) = 0$

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable, $x_0 \in \mathbb{R}^n$, and $H_0 \in \mathbb{R}^{n \times n}$. Suppose that

1. there exists $x \in \mathbb{R}^n$ and $\epsilon > \|x_0 - \bar{x}\|$ such that $f(x) \leq f(x_0)$ whenever $\|x - \bar{x}\| \leq \epsilon$,

2. there is a $\delta > 0$ such that $\delta \|z\|_2^2 \leq z^T \nabla^2 f(x) z$ for all $x \in B(x_0, \epsilon)$,

3. $\nabla^2 f$ is Lipschitz continuous on $\text{cl} B(x; \epsilon)$ with Lipschitz constant $L$,

4. $\theta_0 := L^2 \delta \|x_0 - x\| + M_0 K < 1$ where $M_0 > 0$ satisfies $z^T \nabla^2 f(x) z \leq M_0 \|z\|_2^2$ for all $x \in B(x_0, \epsilon)$ and $K \geq \|\nabla^2 f(x_0) - 1 - H_0\|_y$ with $y_0 = \nabla f(x_0) / \|\nabla f(x_0)\|$.
Newton’s Method for Minimization: \( \nabla f(x) = 0 \)

Let \( f: \mathbb{R}^n \to \mathbb{R} \) be twice continuously differentiable, \( x^0 \in \mathbb{R}^n \), and \( H_0 \in \mathbb{R}^{n \times n} \). Suppose that
Newton’s Method for Minimization: $\nabla f(x) = 0$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $x^0 \in \mathbb{R}^n$, and $H_0 \in \mathbb{R}^{n \times n}$. Suppose that

1. there exists $\bar{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\bar{x}) \leq f(x)$ whenever $\|x - \bar{x}\| \leq \epsilon$, 

Newton’s Method for Minimization: $\nabla f(x) = 0$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $x^0 \in \mathbb{R}^n$, and $H_0 \in \mathbb{R}^{n \times n}$. Suppose that

1. there exists $\bar{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\bar{x}) \leq f(x)$ whenever $\|x - \bar{x}\| \leq \epsilon$,

2. there is a $\delta > 0$ such that $\delta\|z\|_2^2 \leq z^T \nabla^2 f(x)z$ for all $x \in B(\bar{x}, \epsilon)$,
Newton’s Method for Minimization: \( \nabla f(x) = 0 \)

Let \( f : \mathbb{R}^n \to \mathbb{R} \) be twice continuously differentiable, \( x^0 \in \mathbb{R}^n \), and \( H_0 \in \mathbb{R}^{n \times n} \). Suppose that

1. there exists \( \bar{x} \in \mathbb{R}^n \) and \( \epsilon > \|x^0 - \bar{x}\| \) such that \( f(\bar{x}) \leq f(x) \) whenever \( \|x - \bar{x}\| \leq \epsilon \),

2. there is a \( \delta > 0 \) such that \( \delta \|z\|^2_2 \leq z^T \nabla^2 f(x) z \) for all \( x \in B(\bar{x}, \epsilon) \),

3. \( \nabla^2 f \) is Lipschitz continuous on \( \text{cl} B(\bar{x}; \epsilon) \) with Lipschitz constant \( L \), and
Newton’s Method for Minimization: $\nabla f(x) = 0$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $x^0 \in \mathbb{R}^n$, and $H_0 \in \mathbb{R}^{n \times n}$. Suppose that

1. there exists $\bar{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\bar{x}) \leq f(x)$ whenever $\|x - \bar{x}\| \leq \epsilon$,

2. there is a $\delta > 0$ such that $\delta \|z\|_2^2 \leq z^T \nabla^2 f(x) z$ for all $x \in B(\bar{x}, \epsilon)$,

3. $\nabla^2 f$ is Lipschitz continuous on $\text{cl} B(\bar{x}; \epsilon)$ with Lipschitz constant $L$, and

4. $\theta_0 := \frac{L}{2\delta} \|x^0 - \bar{x}\| + M_0 K < 1$ where $M_0 > 0$ satisfies $z^T \nabla^2 f(x) z \leq M_0 \|z\|_2^2$ for all $x \in B(\bar{x}, \epsilon)$ and $K \geq \|(\nabla^2 f(x^0)^{-1} - H_0)y^0\|$ with $y^0 = \nabla f(x^0)/\|\nabla f(x^0)\|$.
Newton’s Method for Minimization: $\nabla f(x) = 0$

Further, suppose that the iteration

$$x^{k+1} := x^k - H_k \nabla f(x^k)$$

is initiated at $x^0$ where the $H_k$’s are chosen to satisfy one of the following conditions:

where for each $k = 1, 2, \ldots$ $y^k := \nabla f(x^k)/\|\nabla f(x^k)\|$. 
Newton’s Method for Minimization: \( \nabla f(x) = 0 \)

Further, suppose that the iteration

\[
x^{k+1} := x^k - H_k \nabla f(x^k)
\]

is initiated at \( x^0 \) where the \( H_k \)'s are chosen to satisfy one of the following conditions:

(i) \( \| (\nabla^2 f(x^k)^{-1} - H_k) y^k \| \leq K \),

where for each \( k = 1, 2, \ldots \) \( y^k := \nabla f(x^k)/\|\nabla f(x^k)\| \).
Newton’s Method for Minimization: $\nabla f(x) = 0$

Further, suppose that the iteration

$$x^{k+1} := x^k - H_k \nabla f(x^k)$$

is initiated at $x^0$ where the $H_k$’s are chosen to satisfy one of the following conditions:

(i) $\|(\nabla^2 f(x^k)^{-1} - H_k)y^k\| \leq K$,

(ii) $\|(\nabla^2 f(x^k)^{-1} - H_k)y^k\| \leq \theta_1^k K$ for some $\theta_1 \in (0, 1)$,

where for each $k = 1, 2, \ldots$ $y^k := \nabla f(x^k)/\|\nabla f(x^k)\|$.
Newton’s Method for Minimization: $\nabla f(x) = 0$

Further, suppose that the iteration

$$x^{k+1} := x^k - H_k \nabla f(x^k)$$

is initiated at $x^0$ where the $H_k$’s are chosen to satisfy one of the following conditions:

(i) $\|(\nabla^2 f(x^k)^{-1} - H_k)y^k\| \leq K$,

(ii) $\|(\nabla^2 f(x^k)^{-1} - H_k)y^k\| \leq \theta_1^k K$ for some $\theta_1 \in (0, 1)$,

(iii) $\|(\nabla^2 f(x^k)^{-1} - H_k)y^k\| \leq \min\{M_2\|x^k - x^{k-1}\|, K\}$, for some $M_2 > 0$, or

where for each $k = 1, 2, \ldots$ $y^k := \nabla f(x^k)/\|\nabla f(x^k)\|$.
Newton’s Method for Minimization: $\nabla f(x) = 0$

Further, suppose that the iteration

$$x^{k+1} := x^k - H_k \nabla f(x^k)$$

is initiated at $x^0$ where the $H_k$’s are chosen to satisfy one of the following conditions:

1. $(\nabla^2 f(x^k)^{-1} - H_k)y^k \leq K$,
2. $(\nabla^2 f(x^k)^{-1} - H_k)y^k \leq \theta_1^k K$ for some $\theta_1 \in (0, 1)$,
3. $(\nabla^2 f(x^k)^{-1} - H_k)y^k \leq \min\{M_2\|x^k - x^{k-1}\|, K\}$, for some $M_2 > 0$, or
4. $(\nabla^2 f(x^k)^{-1} - H_k)y^k \leq \min\{M_2\|\nabla f(x^k)\|, K\}$, for some $M_3 > 0$,

where for each $k = 1, 2, \ldots$ $y^k := \nabla f(x^k)/\|\nabla f(x^k)\|$.
Newton’s Method for Minimization: \( \nabla f(x) = 0 \)

These hypotheses on the accuracy of the approximations \( H_k \) yield the following conclusions about the rate of convergence of the iterates \( x^k \).

(a) If (i) holds, then \( x^k \to \bar{x} \) linearly.

(b) If (ii) holds, then \( x^k \to \bar{x} \) superlinearly.

(c) If (iii) holds, then \( x^\epsilon \to \bar{x} \) two step quadratically.

(d) If (iv) holds, then \( x^k \to \bar{k} \) quadratically.