Rates of Covergence and 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.

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 consider only *quotient rates*, or Q-rates of convergence.

Rates of Convergence

Let $\{x^\nu\} \subset \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ be such that $\bar{x}^\nu \to \bar{x}$.

Rates of Convergence

Let $\{x^\nu\} \subset \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ be such that $\bar{x}^\nu \to \bar{x}$.

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

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

Rates of Convergence

Let $\{x^\nu\} \subset \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ be such that $\bar{x}^\nu \to \bar{x}$.

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.

Rates of Convergence

Let $\{x^\nu\} \subset \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$ be such that $\bar{x}^\nu \to \bar{x}$.

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.

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.

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.

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.

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.

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

Example

Let $f(x) = x^2 + e^x$.
f is a strongly convex function with

$$\begin{aligned} f(x) &= x^2 + e^x \\ f'(x) &= 2x + e^x \\ f''(x) &= 2 + e^x \ > \ 2 \\ f'''(x) &= e^x. \end{aligned}$$

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

Example: Steepest Descent

k	x^k	$f(x^k)$	$f'(x^k)$	s
0	1	.37182818	4.7182818	0
1	0	1	1	0
2	$-.5$	.8565307	-0.3934693	1
3	$-.25$	.8413008	0.2788008	2
4	$-.375$	.8279143	$-.0627107$	3
5	$-.34075$	.8273473	.0297367	5
6	$-.356375$	.8272131	$-.01254$	6
7	$-.3485625$	.8271976	.0085768	7
8	$-.3524688$	.8271848	$-.001987$	8
9	$-.3514922$	.8271841	.0006528	10
10	$-.3517364$	.827184	$-.0000072$	12

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

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

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

x	$f'(x)$
1	4.7182818
0	1
$-1/3$	.0498646
$-.3516893$	.00012
$-.3517337$	.00000000064

Example: Newton's Method

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

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

x	$f'(x)$
1	4.7182818
0	1
$-1/3$	.0498646
$-.3516893$	.00012
$-.3517337$	.00000000064

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 $g : \mathbb{R}^n \to \mathbb{R}^n$, 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 $g : \mathbb{R}^n \to \mathbb{R}^n$ 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. $g(\overline{x}) = 0$,

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. $g(\overline{x}) = 0$,
2. $g'(x)^{-1}$ exists for $x \in B(\overline{x}; \epsilon) := \{x \in \mathbb{R}^n : \|x - \overline{x}\| < \epsilon\}$ with

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

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. $g(\overline{x}) = 0$,
2. $g'(x)^{-1}$ exists for $x \in B(\overline{x}; \epsilon) := \{x \in \mathbb{R}^n : \|x - \overline{x}\| < \epsilon\}$ with
$$\sup\{\|g'(x)^{-1}\| : x \in B(\overline{x}; \epsilon)] \leq M_1$$
3. g' is Lipschitz continuous on $c\ell B(\overline{x}; \epsilon)$ with Lipschitz constant L, and

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. $g(\overline{x}) = 0$,
2. $g'(x)^{-1}$ exists for $x \in B(\overline{x};\epsilon) := \{x \in \mathbb{R}^n : \|x - \overline{x}\| < \epsilon\}$ with

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

3. g' is Lipschitz continuous on $c\ell B(\overline{x};\epsilon)$ with Lipschitz constant L, and
4. $\theta_0 := \frac{LM_1}{2}\|x^0 - \overline{x}\| + M_0K < 1$ where $K \geq \|(g'(x^0)^{-1} - J_0)y^0\|$, $y^0 := g(x^0)/\|g(x^0)\|$, and $M_0 = \max\{\|g'(x)\| : x \in B(\overline{x};\epsilon)\}$.

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;

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 $k = 1, 2, \ldots,$ $y^k := g(x^k)/\|g(x^k)\|$.

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 $\theta_1 \in (0, 1)$,

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

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

(iii) $\|(g'(x^k)^{-1} - J_k)y^k\| \le \min\{M_2\|x^k - x^{k-1}\|, K\}$, 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) $\|(g'(x^k)^{-1} - J_k)y^k\| \leq K \implies x^k \to \bar{x}$ linearly.

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

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: $\nabla f(x) = 0$

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 $\overline{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\overline{x}) \le f(x)$ whenever $\|x - \bar{x}\| \le \epsilon$,

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 $\overline{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\overline{x}) \leq f(x)$ 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 $x \in B(\overline{x}, \epsilon)$,

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 $\overline{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\overline{x}) \leq f(x)$ 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 $x \in B(\overline{x}, \epsilon)$,
3. $\nabla^2 f$ is Lipschitz continuous on $\mathrm{cl} B(\overline{x}; \epsilon)$ 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 $\overline{x} \in \mathbb{R}^n$ and $\epsilon > \|x^0 - \bar{x}\|$ such that $f(\overline{x}) \le f(x)$ whenever $\|x - \bar{x}\| \le \epsilon$,
2. there is a $\delta > 0$ such that $\delta\|z\|_2^2 \le z^T\nabla^2 f(x)z$ for all $x \in B(\overline{x}, \epsilon)$,
3. $\nabla^2 f$ is Lipschitz continuous on $\mathrm{cl} B(\overline{x}; \epsilon)$ with Lipschitz constant L, and
4. $\theta_0 := \frac{L}{2\delta}\|x^0 - \overline{x}\| + M_0 K < 1$ where $M_0 > 0$ satisfies $z^T\nabla^2 f(x)z \le M_0\|z\|_2^2$ for all $x \in B(\overline{x}, \epsilon)$ and $K \ge \|(\nabla^2 f(x^0)^{-1} - H_0)y^0\|$ 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:

(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

(iv) $\|(\nabla^2 f(x^k)^{-1} - H_k)y^k\| \leq \min\{M_2\|\nabla f(x^k)\|, K\}$, 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 $x^k \to \overline{x}$ linearly.

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

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

(d) If (iv) holds, then $x^k \to \overline{k}$ quadradically.