Optimization and Data Science

Lecture 11: Convergence Measures for Iterative Algorithms

Prof. Dr. Thomas Slawig

Kiel University - CAU Kiel
Dep. of Computer Science

Summer 2020

Contents

Contents

Motivation

- General descent methods and the gradient method generate sequences $(x_k)_k$ of iterates.
- We want to make sure that these sequences converge to a minimizer x^* ...
- ... and we want to measure the convergence speed.
- The convergence speed of different sequences converging all to the same point can be quite different.
- The convergence speed of iterative (optimization) algorithms is a main quality criterion, ...
- ... since it indicates how many iteration steps are necessary to achieve a prescribed, desired accuracy.
- The second criterion is the necessary computational effort per iteration step.

Motivation

- Consider the following sequences:

$$(x_k)_{k\in\mathbb{N}} = \left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots\right), \text{ i.e., } x_k = \left(\frac{1}{2}\right)^k$$

$$(x_k)_{k\in\mathbb{N}} = \left(1, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right), \text{ i.e., } x_0 = 1, x_k = \frac{1}{k}, k \geq 1,$$

$$(x_k)_{k\in\mathbb{N}} = \left(1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \dots\right), \text{ i.e., } x_k = \frac{1}{k!}$$

$$(x_k)_{k\in\mathbb{N}} = \left(\frac{3}{2}, \frac{1}{2}s, \frac{3}{2}s^2, \frac{1}{2}s^3, \dots\right), \text{ i.e., } x_k = s^k\left(1 + \frac{1}{2}(-1)^k\right), \quad s \in (0,1)$$

$$(x_k)_{k\in\mathbb{N}} = \left(1, 1, 1, 2, \frac{2}{9}, \frac{8}{9}, \frac{8}{225}, \dots\right), \text{ i.e., } x_0 = x_1 = 1, x_{k+1} = \begin{cases} kx_k & \text{for even } k \geq 2 \\ \frac{1}{k^2}x_k & \text{for odd } k \end{cases}$$

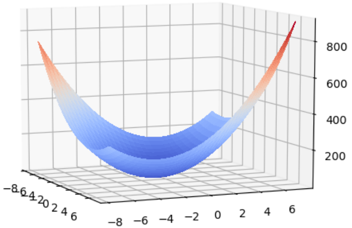
- All converge to $x^* = 0$, but with quite different speed.

Contents

Convergence speed for quadratic functions

- General quadratic function $f : \mathbb{R}^n \to \mathbb{R}$:

$$f(x) = \frac{1}{2}x^\top Ax + b^\top x + c, \quad A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n, c \in \mathbb{R}.$$



- Exact step-size:

$$\rho_k = -\frac{(Ax_k + b)^\top d_k}{d_k^\top A d_k}.$$

- Example: diagonal matrix

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2a \end{pmatrix}$$

- Convergence:

$$\|x_{k+1} - x^*\|_2 = \frac{a-1}{a+1}\|x_k - x^*\|_2 \quad \text{for all } k \in \mathbb{N}$$

- ***Q*-factor** (quotient factor).
- $a = 1 : Q = 0 \rightsquigarrow$ fast convergence, $a \gg 1 : Q \approx 1 \rightsquigarrow$ slow.

Contents

Measuring convergence speed

The first way to describe or measure the convergence speed of a sequence $(x_k)_k$ (if the limit x^* is known):

- We measure the factor by which the distance to the limit is reduced in one step, i.e., for each k the number $q_k \geq 0$ satisfying

$$\|x_{k+1} - x^*\| \leq q_k \|x_k - x^*\|$$

or equivalently

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq q_k \quad (\text{if } x_k \neq x^*).$$

- This generates a sequence $(q_k)_k$.

Q-factor

Definition

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence with $x_k \to x^*$ (and $x_k \neq x^*$ for all k). If the sequence $(q_k)_k$ satisfying

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq q_k \quad (\text{if } x_k \neq x^*).$$

- is constant $= q$
- or converges to some $q = \lim_{k\to\infty} q_k$,

we call q the Q-**factor** (Quotient factor) of the sequence $(x_k)_{k\in\mathbb{N}}$.

Example 1 Q-factor

We consider the sequence $(x_k)_{k\in\mathbb{N}}$ given as

$$(x_k)_k = \left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots\right),$$

i.e.,

$$x_k = \left(\frac{1}{2}\right)^k \to x^* = 0.$$

Compute the values q_k:

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{\left(\frac{1}{2}\right)^{k+1}}{\left(\frac{1}{2}\right)^k} = q_k = \frac{1}{2} \text{ for all } k \in \mathbb{N},$$

and thus the Q-factor is $\frac{1}{2}$.

Example 2 Q-factor

We consider the sequence $(x_k)_{k\in\mathbb{N}}$ given as

$$(x_k)_k = \left(1, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right), \text{ i.e., } x_0 = 1, k \geq 1 : x_k = \frac{1}{k} \to x^* = 0.$$

Compute the values q_k and the limit of the sequence $(q_k)_k$:

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k}{k+1} = q_k \quad \text{for } k \geq 1,$$

This gives

$$\lim_{k\to\infty} q_k = 1,$$

thus the Q-factor is 1.

Example 3 Q-factor

We consider the sequence $(x_k)_{k\in\mathbb{N}}$ given as

$$(x_k)_k = \left(1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \dots\right),$$

i.e.,

$$x_k = \frac{1}{k!} \to x^* = 0.$$

Compute the values q_k and the limit of the sequence $(q_k)_k$:

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k!}{(k+1)!} = \frac{1}{k+1} = q_k,$$

This gives

$$\lim_{k\to\infty} q_k = 0,$$

thus the Q-factor is 0.

Example 4 Q-factor: alternating sequence

Let $s \in (0,1)$ be arbitrary. We consider

$$(x_k)_{k\in\mathbb{N}} = \left(\frac{3}{2}, \frac{1}{2}s, \frac{3}{2}s^2, \frac{1}{2}s^3, \dots\right), \text{ i.e., } x_k = s^k\left(1 + \frac{1}{2}(-1)^k\right) \to 0.$$

Compute the values q_k. What do you observe?

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \left\{\begin{array}{lll} \frac{\frac{1}{2}s^{k+1}}{\frac{3}{2}s^k} = & \frac{s}{3} & \text{for even } k \\ \frac{\frac{3}{2}s^{k+1}}{\frac{1}{2}s^k} = & 3s & \text{for odd } k \end{array}\right\} = q_k.$$

- The sequence $(q_k)_k$ does not converge. Our definition is not applicable.
- $(q_k)_k$ consists of two subsequences, each of which converges to an **accumulation point**.
- The bigger one determines the convergence speed of $(x_k)_k$.

Example 5 Q-factor: alternating sequence

We consider

$$(x_k)_{k\in\mathbb{N}} = \left(1, 1, 1, 2, \frac{2}{9}, \frac{8}{9}, \frac{8}{225}, \ldots\right), \text{ i.e., } x_0 = x_1 = 1, x_{k+1} = \left\{ \begin{array}{ll} kx_k & \text{even } k \geq 2 \\ \frac{1}{k^2}x_k & \text{odd } k \end{array} \right\} \to 0.$$

Compute the values q_k. What do you observe?

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{x_{k+1}}{x_k} = \left\{ \begin{array}{ll} k & \text{for even } k \geq 2 \\ \dfrac{1}{k^2} & \text{for odd } k \end{array} \right\} = q_k.$$

- The sequence $(q_k)_k$ does not converge. Our definition is not applicable.
- $(q_k)_k$ consists of two subsequences, one converges to zero, the other one does not converge.

Contents

Accumulation points

Definition

For a sequence $(x_k)_k$ we define the **limit superior** or **biggest accumulation point** as

$$\limsup_{k\to\infty} x_k := \lim_{k\to\infty} \sup\{x_\ell : \ell \geq k\},$$

where

$$\left.\sup\{x_\ell : \ell \geq k\} := \begin{cases} \min\{M \in \mathbb{R} : x_\ell \leq M \text{ for all } \ell \geq k\}, & \text{if the minimum exists} \\ \infty, & \text{if not} \end{cases}\right\} k \in \mathbb{N},$$

is the **lowest upper bound** of the considered subsequence $\{x_\ell : \ell \geq k\}$.

- For a converging sequence, the limit superior equals the limit.
- It differs only for non-converging sequences, ...
- ... and it is only important for sequences with two (or more) converging subsequences.
- The smallest accumulation point is the **limit inferior** that we do not need here.

Example: Accumulation points (1)

- Consider the sequence:

$$x_k = (-1)^k.$$

- Obviously it is not convergent,
- ... but has two accumulation points.
- The subsequence (with odd k) converges to/is constant -1.
- The subsequence (with even k) converges to/is constant 1.
- We have

$$\sup\{x_\ell : \ell \geq k\} = \min\{M \in \mathbb{R} : x_\ell \leq M \text{ for all } \ell \geq k\} = 1 \text{ for all } k.$$

Thus the biggest accumulation point or limit superior is

$$\limsup_{k\to\infty} x_k = \lim_{k\to\infty} \sup\{x_\ell : \ell \geq k\} = 1.$$

Example: Accumulation points (2)

- Consider the sequence:

$$x_k = (-1)^k + \frac{1}{k}.$$

- Obviously it is not convergent.
- We have

$$\sup\{x_\ell : \ell \geq k\} = \min\{M \in \mathbb{R} : x_\ell \leq M \text{ for all } \ell \geq k\} = 1 + \frac{1}{k} \text{ for all } k.$$

Thus the limit superior is

$$\limsup_{k\to\infty} x_k = \lim_{k\to\infty} \sup\{x_\ell : \ell \geq k\} = \lim_{k\to\infty} \left(1 + \frac{1}{k}\right) = 1.$$

- The biggest accumulation point is again 1.

General definition of the Q-factor

Definition

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence with $x_k \to x^*$ (and $x_k \neq x^*$ for all k). We call

$$Q((x_k)_{k\in\mathbb{N}}) := \limsup_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$$

the Q-**factor** (Quotient factor) of the sequence .

In other words:

- For each k, we find the smallest number $q_k \in \mathbb{R}_{\geq 0}$ that satisfies

$$\|x_{k+1} - x^*\| \leq q_k \|x_k - x^*\|.$$

- Then, the Q-factor is the biggest accumulation point of the sequence $(q_k)_{k\in\mathbb{N}}$.
- It describes (an upper bound for) the convergence speed of the sequence $(x_k)_{k\in\mathbb{N}}$.

Example 4 Q-factor: alternating sequence revisited

Let $s \in (0,1)$ be arbitrary. We consider

$$(x_k)_{k\in\mathbb{N}} = \left(\frac{3}{2}, \frac{1}{2}s, \frac{3}{2}s^2, \frac{1}{2}s^3, \dots\right), \text{ i.e., } x_k = s^k\left(1 + \frac{1}{2}(-1)^k\right) \to 0.$$

We computed

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \left\{ \begin{array}{ccll} \frac{\frac{1}{2}s^{k+1}}{\frac{3}{2}s^k} & = & \frac{s}{3} & \text{for even } k \\ \frac{\frac{3}{2}s^{k+1}}{\frac{1}{2}s^k} & = & 3s & \text{for odd } k \end{array} \right\} = q_k.$$

What is the biggest accumulation point, i.e., the Q-factor?

$$\limsup_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 3s.$$

Example 5: Q-factor for alternating sequence

We consider

$$(x_k)_{k\in\mathbb{N}} = \left(1, 1, 1, 2, \frac{2}{9}, \frac{8}{9}, \frac{8}{225}, \ldots\right), \text{ i.e., } x_0 = x_1 = 1, x_{k+1} = \left\{ \begin{array}{ll} kx_k, & \text{even } k \geq 2 \\ \frac{1}{k^2}x_k, & \text{odd } k \end{array} \right\} \to 0.$$

We computed

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \left\{ \begin{array}{ll} k & \text{for even } k \geq 2 \\ \dfrac{1}{k^2} & \text{for odd } k \end{array} \right\} = q_k.$$

The subsequence for odd k converges to the accumulation point $\bar{q} = 0$.
The other subsequence, for even k, does not converge.
Thus, the Q-factor is

$$\limsup_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \limsup_{k\to\infty} q_k = \lim_{k\to\infty} k = \infty.$$

Contents

Measuring average convergence speed

If $(q_k)_k$ has different accumulation points, we may measure the average convergence speed:

- If we have as above

$$\|x_{k+1} - x^*\| \leq q_k \|x_k - x^*\|,$$

- ... we get for the first k steps

$$\|x_k - x^*\| \leq q_{k-1} \cdots q_1 q_0 \|x_0 - x^*\|.$$

- We compute the geometric mean of the q_k, i.e.,

$$r_k = \sqrt[k]{q_{k-1} \cdots q_1 q_0}.$$

or

$$\sqrt[k]{\frac{\|x_k - x^*\|}{\|x_0 - x^*\|}} \leq r_k.$$

R-factor

Definition

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence with $x_k \to x^*$ (and $x_0 \neq x^*$). We call

$$R((x_k)_{k\in\mathbb{N}}) := \limsup_{k\to\infty} \sqrt[k]{\frac{\|x_k - x^*\|}{\|x_0 - x^*\|}}$$

the R-**factor** (Root factor) of the sequence $(x_k)_{k\in\mathbb{N}}$.

- The R-factor describes the <u>average</u> convergence speed of the sequence $(x_k)_{k\in\mathbb{N}}$.
- If $(q_k)_k$ is constant, then R-factor $=$ Q-factor.
- If $(q_k)_k$ converges to some $q \in \mathbb{R}$, then R-factor $=$ Q-factor.
- In general: R-factor $\leq$ Q-factor.

Example 1: $(q_k)_k$ constant $\rightsquigarrow$ R-factor $=$ Q-factor

We consider the sequence $(x_k)_{k\in\mathbb{N}}$ given as

$$(x_k)_{k\in\mathbb{N}} = \left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots\right), \text{ i.e., } x_k = \left(\frac{1}{2}\right)^k \to x^* = 0.$$

We had

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = q_k = \frac{1}{2} \text{ for all } k \in \mathbb{N},$$

i.e., $Q((x_k)_{k\in\mathbb{N}}) = \frac{1}{2}$. We get

$$\sqrt[k]{\frac{\|x_k - x^*\|}{\|x_0 - x^*\|}} = \sqrt[k]{\frac{\left(\frac{1}{2}\right)^k}{1}} = \sqrt[k]{\left(\frac{1}{2}\right)^k} = \frac{1}{2} \text{ for all } k \in \mathbb{N}$$

and thus

$$R((x_k)_{k\in\mathbb{N}}) = \frac{1}{2}.$$

Example 2: $(q_k)_k$ convergent $\rightsquigarrow$ R-factor $=$ Q-factor

For the sequence $(x_k)_k = \left(\frac{1}{k}\right)_k, k \geq 1$, we had

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k}{k+1} = q_k \to 1,$$

Thus, the Q-factor is 1. Since $(q_k)_k$ converges, the R-factor is the same.
We can also compute it directly:

$$\sqrt[k]{\frac{\|x_k - x^*\|}{\|x_0 - x^*\|}} = \sqrt[k]{\frac{\frac{1}{k}}{1}} = \sqrt[k]{\frac{1}{k}} = \frac{1}{\sqrt[k]{k}} \text{ for all } k \geq 1.$$

Since

$$\sqrt[k]{k} = k^{\frac{1}{k}} \quad \Leftrightarrow \quad \log\left(k^{\frac{1}{k}}\right) = \frac{1}{k}\log k = \frac{\log k}{k} \to 0$$

we get

$$\lim_{k\to\infty} \sqrt[k]{k} = e^0 = 1 \quad \Rightarrow \quad \lim_{k\to\infty} \frac{1}{\sqrt[k]{k}} = 1.$$

Example 3: $(q_k)_k$ convergent $\rightsquigarrow$ R-factor $=$ Q-factor

We consider $(x_k)_k = \left(\frac{1}{k!}\right)_k$. We had

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k!}{(k+1)!} = \frac{1}{k+1} = q_k \to 0.$$

The Q-factor was 0. Since $(q_k)_k$ converges, the R-factor is the same. We can also compute it directly:

$$\sqrt[k]{\frac{\|x_k - x^*\|}{\|x_0 - x^*\|}} = \sqrt[k]{\frac{\frac{1}{k!}}{1}} = \sqrt[k]{\frac{1}{k!}} = \frac{1}{\sqrt[k]{k!}} \to 0$$

since it can be shown that

$$\lim_{k\to\infty} \sqrt[k]{k!} = \infty.$$

Example 4: R-factor for alternating sequence

We consider again $(x_k)_k = \left(s^k\left(1+\frac{1}{2}(-1)^k\right)\right)_k$ with $s \in (0,1)$. We had

$$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = \left\{\begin{array}{lll} \frac{\frac{1}{2}s^{k+1}}{\frac{3}{2}s^k} = & \frac{s}{3} & \text{for even } k \\ \frac{\frac{3}{2}s^{k+1}}{\frac{1}{2}s^k} = & 3s & \text{for odd } k \end{array}\right\} = q_k.$$

The Q-factor was the biggest accumulation point: $3s$. Compute the R-factor:

$$\sqrt[k]{\frac{\|x_k-x^*\|}{\|x_0-x^*\|}} = \left\{\begin{array}{lll} \sqrt[k]{\frac{\frac{3}{2}s^k}{\frac{3}{2}}} = & s & \text{for even } k, \\ \sqrt[k]{\frac{\frac{1}{2}s^k}{\frac{3}{2}}} = & \underbrace{\sqrt[k]{\frac{1}{3}}}_{\to 1}\, s & \text{for odd } k. \end{array}\right.$$

We thus get that the R-factor is s.

Example 5: R-factor for sequence with $Q = \infty$

We consider

$$(x_k)_{k\in\mathbb{N}} = \left(1, 1, 1, 2, \frac{2}{9}, \frac{8}{9}, \frac{8}{225}, \dots\right), \text{ i.e., } x_0 = x_1 = 1, x_{k+1} = \left\{ \begin{array}{ll} kx_k, & \text{even } k \geq 2 \\ \frac{1}{k^2}x_k, & \text{odd } k \end{array} \right\} \to 0.$$

The Q-factor was ∞. We computed

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \begin{cases} k & \text{for even } k \geq 2 \\ \dfrac{1}{k^2} & \text{for odd } k \end{cases}$$

We get

$$\frac{\|x_{k+2} - x^*\|}{\|x_k - x^*\|} = \frac{\|x_{k+2} - x^*\|}{\|x_{k+1} - x^*\|} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \left\{ \begin{array}{l} \dfrac{k}{(k+1)^2} \leq \dfrac{k}{k^2} = \dfrac{1}{k}, \text{ even } k \geq 2 \\ \dfrac{k+1}{k^2} \leq \dfrac{2k}{k^2} = \dfrac{2}{k}, \text{ odd } k \end{array} \right\} \leq \frac{2}{k}, k \geq 2.$$

Example 5: R-factor for sequence with $Q = \infty$

We have $x_0 = x_1 = x_2 = 1$ and

$$\frac{\|x_{k+2} - x^*\|}{\|x_k - x^*\|} \le \frac{2}{k} \text{ for all } k \ge 2.$$

Thus,

$$\frac{\|x_k - x^*\|}{\|x_0 - x^*\|} = \frac{\|x_k - x^*\|}{\|x_{k-2} - x^*\|} \cdots \frac{\|x_4 - x^*\|}{\|x_2 - x^*\|} \frac{\|x_2 - x^*\|}{\|x_0 - x^*\|} \le \underbrace{\frac{2}{k-2}}_{<1} \underbrace{\frac{2}{k-4}}_{<1} \cdots \underbrace{\frac{2}{4}}_{<1} \underbrace{\frac{2}{2} \cdot 1}_{=1} < 1$$

Since

$$\lim_{k\to\infty} \sqrt[k]{c} = 1 \text{ for all } 0 < c < 1,$$

the R-factor is 1.

Comparison: Q- and R-factors for the examples

$(x_k)_{k\in\mathbb{N}}$	$\left(\frac{1}{2}\right)^k$	$\frac{1}{k}$	$\frac{1}{k!}$	$s^k\left(1+\frac{1}{2}(-1)^k\right),\ s\in(0,1)$	$x_{k+1}=\begin{cases} kx_k, & \text{even } k\geq 2 \\ \frac{1}{k^2}x_k, & \text{odd } k \end{cases}$
Q-factor	$\frac{1}{2}$	1	0	$3s$	∞
R-factor	$\frac{1}{2}$	1	0	s	1

- Difference only if the convergence behavior is not uniform (as in the 4th example).
- This often occurs for sequences of iterates in optimization.

Contents

Convergence rates: Linear and superlinear convergence

This motivates to define the following notions:

Definition

A converging sequence $(x_k)_k$ is called

- **Q-linearly convergent** if $Q((x_k)_k) < 1$.
- **R-linearly convergent** if $R((x_k)_k) < 1$.
- **Q-sublinearly convergent** if $Q((x_k)_k) \geq 1$.
- **R-sublinearly convergent** if $R((x_k)_k) = 1$.
- **Q-superlinearly convergent** if $Q((x_k)_k) = 0$.
- **R-superlinearly convergent** if $R((x_k)_k) = 0$.

Comparison: Convergence rates for the examples

$(x_k)_{k\in\mathbb{N}}$	$\left(\frac{1}{2}\right)^k$	$\frac{1}{k}$	$\frac{1}{k!}$	$s^k\left(1+\frac{1}{2}(-1)^k\right), s\in(0,1)$	$x_{k+1}=\begin{cases} kx_k, & \text{even } k\geq 2 \\ \frac{1}{k^2}x_k, & \text{odd } k\end{cases}$
Q-factor	$\frac{1}{2}$	1	0	$3s$	∞
Q-conv.	linear	sub-linear	super-linear	linear for $s\in(0,\frac{1}{3})$ sublinear for $s\in[\frac{1}{3},1)$	sublinear
R-factor	$\frac{1}{2}$	1	0	s	1
R-conv.	linear	sub-linear	super-linear	linear for all $s\in(0,1)$	sublinear

What is important

- One important quality criterion of iterative algorithms is the convergence speed.
- We measure it with Q- and R-factors.
- The Q-factor describes the maximum of the reduction in the distance between iterate and the limit point.
- The R-factor describes the geometric average of this reduction.
- Using these factors, we define linear, superlinear and sublinear convergence.