Convergence rates in convex optimization
Going beyond the worst-case analysis

Guillaume Garrigos

Laboratory for Computational and Statistical Learning
Istituto Italiano di Tecnologia & Massachusetts Institute of Technology
Genoa, Italy

ICCOPT 2016

Introduction

Setting: X Hilbert space, $f \in \Gamma_0(X)$, argmin $f \neq \emptyset$.

As optimizers, we often face the same questions concerning the convergence of an algorithm:

- **(Qualitative result)** For the iterates: weak, strong convergence?
- **(Quantitative result)** For the values and/or the iterates: super linear, linear $O(\gamma^n)$, polynomial $O(n^{-\gamma})$ rates?

Setting: X Hilbert space, $f \in \Gamma_0(X)$, argmin $f \neq \emptyset$.

As optimizers, we often face the same questions concerning the convergence of an algorithm:

- **(Qualitative result)** For the iterates: weak, strong convergence?
- **(Quantitative result)** For the values and/or the iterates: super linear, linear $O\left(\gamma^n\right)$, polynomial $O\left(n^{-\gamma}\right)$ rates?

It depends on the algorithm and the assumptions made on f.

Introduction

Setting: X Hilbert space, $f \in \Gamma_0(X)$, argmin $f \neq \emptyset$.

As optimizers, we often face the same questions concerning the convergence of an algorithm:

- **(Qualitative result)** For the iterates: weak, strong convergence?
- **(Quantitative result)** For the values and/or the iterates: super linear, linear $O(\gamma^n)$, polynomial $O(n^{-\gamma})$ rates?

It depends on the algorithm and the assumptions made on f.

Here we will essentially consider first order descent methods, and more simply the forward-backward method.

Introduction

Setting: X Hilbert space, $f \in \Gamma_0(X)$, argmin $f \neq \emptyset$.

As optimizers, we often face the same questions concerning the convergence of an algorithm:

- **(Qualitative result)** For the iterates: weak, strong convergence?
- **(Quantitative result)** For the values and/or the iterates: super linear, linear $O(\gamma^n)$, polynomial $O(n^{-\gamma})$ rates?

It depends on the algorithm and the assumptions made on f.

Here we will essentially consider first order descent methods, and more simply the forward-backward method.

Contents

Classic convergence results

Assume $\operatorname{argmin} f \neq \emptyset$ and $(x_n)_{n\in\mathbb{N}}$ be generated by forward-backward.

Theorem (general convex case)

x_n weakly converges to $x_\infty \in \operatorname{argmin} f$, and $f(x_n) - \inf f = O\left(\frac{1}{n}\right)$.

Classic convergence results

Assume $\operatorname{argmin} f \neq \emptyset$ and $(x_n)_{n\in\mathbb{N}}$ be generated by forward-backward.

Theorem (general convex case)

x_n weakly converges to $x_\infty \in \operatorname{argmin} f$, and $f(x_n) - \inf f = O\left(\frac{1}{n}\right)$.

Theorem (strongly convex case)

Assume that f is strongly convex. Then x_n strongly converges to $x_\infty \in \operatorname{argmin} f$, and both iterates and values converge linearly.

Classic convergence results

Assume $\operatorname{argmin} f \neq \emptyset$ and $(x_n)_{n\in\mathbb{N}}$ be generated by forward-backward.

function	convex		s. convex
iterates	weak cv		linear
values	n^{-1}		linear

Classic convergence results

Assume argmin $f \neq \emptyset$ and $(x_n)_{n \in \mathbb{N}}$ be generated by forward-backward.

function	convex	?	s. convex
iterates	weak cv		linear
values	n^{-1}		linear

Known examples

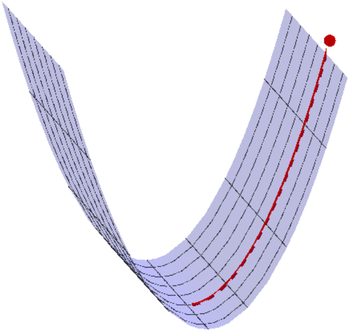
$A \in L(X, Y)$, $y \in Y$.

- $f(x) = \frac{1}{2}\|Ax - y\|^2$, $x_{n+1} = x_n - \tau A^*(Ax_n - y)$
  - If $R(A)$ is closed, linear convergence.

Known examples

$A \in L(X, Y)$, $y \in Y$.

- $f(x) = \frac{1}{2}\|Ax - y\|^2$, $x_{n+1} = x_n - \tau A^*(Ax_n - y)$
  - If $R(A)$ is closed, linear convergence.



Known examples

$A \in L(X, Y)$, $y \in Y$.

- $f(x) = \frac{1}{2}\|Ax - y\|^2$, $x_{n+1} = x_n - \tau A^*(Ax_n - y)$
  - If $R(A)$ is closed, linear convergence.
  - Else, strong convergence for iterates, arbitrarily slow.

Known examples

$A \in L(X, Y)$, $y \in Y$.

- $f(x) = \frac{1}{2}\|Ax - y\|^2$, $x_{n+1} = x_n - \tau A^*(Ax_n - y)$
  - If $R(A)$ is closed, linear convergence.
  - Else, strong convergence for iterates, arbitrarily slow.

- $f(x) = \alpha\|x\|_1 + \frac{1}{2}\|Ax - y\|^2$, $x_{n+1} = \mathcal{S}_{\alpha\tau}\left(x_n - \tau A^*(Ax_n - y)\right)$
  - In $X = \ell^2(\mathbb{N})$, ISTA converges strongly[1].
  - In $X = \mathbb{R}^N$, the convergence is linear.[2]
  - In $X = \ell^2(\mathbb{N})$, linear rates can also be obtained under some conditions[3].

---

[1] Daubechies, Defrise, DeMol - 2004

[2] Liang, Fadili, Peyré - 2014 & Bolte, Nguyen, Peypouquet, Suter - 2015

[3] Bredies, Lorenz - 2008

Contents

The Lojasiewicz inequality

Definition (Lojasiewicz, 1965)

Let $x^* \in \text{argmin } f$ and $\theta \in]0, 1/2]$. We say that f has the Lojasiewicz property at x^* with exponent θ if $\exists C > 0$, $\delta > 0$ s.t.

$$x \in x^* + \delta\mathbb{B}_X \text{ and } f(x) > \inf f \Rightarrow (f(x) - \inf f)^{1-\theta} \leq C\|\partial f(x)\|_-$$

[4] Bolte, Nguyen, Peypouquet, Suter - 2015

The Lojasiewicz inequality

Definition (Lojasiewicz, 1965)

Let $x^* \in \operatorname{argmin} f$ and $\theta \in]0, 1/2]$. We say that f has the Lojasiewicz property at x^* with exponent θ if $\exists C > 0$, $\delta > 0$ s.t.

$$x \in x^* + \delta\mathbb{B}_X \text{ and } f(x) > \inf f \Rightarrow (f(x) - \inf f)^{1-\theta} \leq C\|\partial f(x)\|_-$$

In our convex case, equivalent to an error bound[4]

$$x \in x^* + \delta\mathbb{B}_X \text{ and } f(x) > \inf f \Rightarrow C' \operatorname{dist}(x, \operatorname{argmin} f)^{1/\theta} \leq f(x) - \inf f$$

[4] Bolte, Nguyen, Peypouquet, Suter - 2015

The Lojasiewicz inequality

Definition (Lojasiewicz, 1965)

Let $x^* \in \operatorname{argmin} f$ and $\theta \in]0, 1/2]$. We say that f has the Lojasiewicz property at x^* with exponent θ if $\exists C > 0$, $\delta > 0$ s.t.

$$x \in x^* + \delta\mathbb{B}_X \text{ and } f(x) > \inf f \Rightarrow (f(x) - \inf f)^{1-\theta} \leq C\|\partial f(x)\|_-$$

In our convex case, equivalent to an error bound[4]

$$x \in x^* + \delta\mathbb{B}_X \text{ and } f(x) > \inf f \Rightarrow C' \operatorname{dist}(x, \operatorname{argmin} f)^{1/\theta} \leq f(x) - \inf f$$

The Lojasiewicz exponent θ governs the local geometry of f, and then the rates of convergence.

[4] Bolte, Nguyen, Peypouquet, Suter - 2015

The Lojasiewicz inequality: Examples

$$C\ \mathrm{dist}\,(x, \text{argmin } f)^{1/\theta} \leq f(x) - \inf f$$

- strongly convex functions are Lojasiewicz functions for $\theta = 1/2$

The Lojasiewicz inequality: Examples

$$C \operatorname{dist}(x, \text{argmin } f)^{1/\theta} \leq f(x) - \inf f$$

- strongly convex functions are Lojasiewicz functions for $\theta = 1/2$
- $f(x) = \frac{1}{2}\|Ax - y\|^2$
  - If $R(A)$ is closed, f is Lojasiewicz with $\theta = 1/2$.
  - Else, complicated (?).

The Lojasiewicz inequality: Examples

$$C \,\mathrm{dist}\,(x, \operatorname{argmin} f)^{1/\theta} \leq f(x) - \inf f$$

- strongly convex functions are Lojasiewicz functions for $\theta = 1/2$
- $f(x) = \frac{1}{2}\|Ax - y\|^2$
  - If $R(A)$ is closed, f is Lojasiewicz with $\theta = 1/2$.
  - Else, complicated (?).
- $f(x) = \alpha\|x\|_1 + \frac{1}{2}\|Ax - y\|^2$: In $X = \mathbb{R}^N$, it is[1] a Lojasiewicz function with $\theta = 1/2$.

[1] Bolte, Nguyen, Peypouquet, Suter - 2015

The Lojasiewicz inequality: Examples

$$C \operatorname{dist}(x, \text{argmin } f)^{1/\theta} \leq f(x) - \inf f$$

- strongly convex functions are Lojasiewicz functions for $\theta = 1/2$
- $f(x) = \frac{1}{2}\|Ax - y\|^2$
  - If $R(A)$ is closed, f is Lojasiewicz with $\theta = 1/2$.
  - Else, complicated (?).
- $f(x) = \alpha\|x\|_1 + \frac{1}{2}\|Ax - y\|^2$: In $X = \mathbb{R}^N$, it is[1] a Lojasiewicz function with $\theta = 1/2$.
- semi-algebraic functions are Lojasiewicz functions[2].

[1] Bolte, Nguyen, Peypouquet, Suter - 2015

[2] Bolte, Daniilidis, Lewis, Shiota - 2007

The Lojasiewicz inequality: Convergence result

Theorem (Frankel, Garrigos, Peypouquet - 2014)

Let $(x_n)_{n\in\mathbb{N}}$ be generated by the Forward-Backward method, and assume that $(x_n)_{n\in\mathbb{N}} \subset x^* + \delta\mathbb{B}_X$ where the Lojasiewicz inequality holds. Then x_n converges strongly to a minimizer $x^\dagger$ of f.

The Lojasiewicz inequality: Convergence result

Theorem (Frankel, Garrigos, Peypouquet - 2014)

Let $(x_n)_{n\in\mathbb{N}}$ be generated by the Forward-Backward method, and assume that $(x_n)_{n\in\mathbb{N}} \subset x^* + \delta\mathbb{B}_X$ where the Lojasiewicz inequality holds. Then x_n converges strongly to a minimizer $x^\dagger$ of f.
Moreover, depending on the Lojasiewicz exponent θ:

1. If $\theta = 1/2$, then both iterates $\|x_n - x^\dagger\|$ and values $f(x_n) - \inf f$ converge linearly.
2. If $\theta \in]0, 1/2[$, then iterates and values converge polynomially:
   - $\|x_n - x^\dagger\| = O\left(n^{\frac{-\theta}{1-2\theta}}\right)$
   - $f(x_n) - \inf f = O\left(n^{\frac{-1}{1-2\theta}}\right)$

The Lojasiewicz inequality: Convergence result

Theorem (Frankel, Garrigos, Peypouquet - 2014)

1. If $\theta = 1/2$, then both iterates $\|x_n - x^\dagger\|$ and values $f(x_n) - \inf f$ converge linearly.
2. If $\theta \in]0, 1/2[$, then iterates and values converge polynomially:
   - $\|x_n - x^\dagger\| = O\left(n^{\frac{-\theta}{1-2\theta}}\right)$
   - $f(x_n) - \inf f = O\left(n^{\frac{-1}{1-2\theta}}\right)$

Some remarks on the convergence result:

The Lojasiewicz inequality: Convergence result

Theorem (Frankel, Garrigos, Peypouquet - 2014)

1. If $\theta = 1/2$, then both iterates $\|x_n - x^\dagger\|$ and values $f(x_n) - \inf f$ converge linearly.
2. If $\theta \in]0, 1/2[$, then iterates and values converge polynomially:
   - $\|x_n - x^\dagger\| = O\left(n^{\frac{-\theta}{1-2\theta}}\right)$
   - $f(x_n) - \inf f = O\left(n^{\frac{-1}{1-2\theta}}\right)$

Some remarks on the convergence result:

- Holds for general first-order descent methods:
  1. **(descent)** $a\|x_{n+1} - x_n\|^2 \leq f(x_n) - f(x_{n+1})$
  2. **(1st order)** $b\|\partial f(x_{n+1})\|_{\_} \leq \|x_{n+1} - x_n\|$

  Allows even more structured methods (decomposition by blocs), or variants (variable metric, inexact computations)

The Lojasiewicz inequality: Convergence result

Theorem (Frankel, Garrigos, Peypouquet - 2014)

1. If $\theta = 1/2$, then both iterates $\|x_n - x^\dagger\|$ and values $f(x_n) - \inf f$ converge linearly.
2. If $\theta \in]0, 1/2[$, then iterates and values converge polynomially:
   - $\|x_n - x^\dagger\| = O\left(n^{\frac{-\theta}{1-2\theta}}\right)$
   - $f(x_n) - \inf f = O\left(n^{\frac{-1}{1-2\theta}}\right)$

Some remarks on the convergence result:

- Holds for general first-order descent methods:
  1. **(descent)** $a\|x_{n+1} - x_n\|^2 \leq f(x_n) - f(x_{n+1})$
  2. **(1st order)** $b\|\partial f(x_{n+1})\|_{\_} \leq \|x_{n+1} - x_n\|$

  Allows even more structured methods (decomposition by blocs), or variants (variable metric, inexact computations)
- Holds also in the non-convex setting.

The Lojasiewicz inequality: Convergence result

Theorem (Frankel, Garrigos, Peypouquet - 2014)

1. If $\theta = 1/2$, then both iterates $\|x_n - x^\dagger\|$ and values $f(x_n) - \inf f$ converge linearly.
2. If $\theta \in]0, 1/2[$, then iterates and values converge polynomially:
   - $\|x_n - x^\dagger\| = O\left(n^{\frac{-\theta}{1-2\theta}}\right)$
   - $f(x_n) - \inf f = O\left(n^{\frac{-1}{1-2\theta}}\right)$

Some remarks on the convergence result:

- Holds for general first-order descent methods:
  1. **(descent)** $a\|x_{n+1} - x_n\|^2 \leq f(x_n) - f(x_{n+1})$
  2. **(1st order)** $b\|\partial f(x_{n+1})\|_{\_} \leq \|x_{n+1} - x_n\|$

  Allows even more structured methods (decomposition by blocs), or variants (variable metric, inexact computations)
- Holds also in the non-convex setting.
- Works for other modulus than $t \mapsto t^\theta$ (Kurdyka-Lojasiewicz).

Updated results

Assume argmin $f \neq \emptyset$ and $(x_n)_{n\in\mathbb{N}}$ be generated by a first-order descent method.

function	convex	$\theta \in]0, \frac{1}{2}[$	$\theta = \frac{1}{2}$
iterates	weak cv	$n^{\frac{-\theta}{1-2\theta}}$	linear
values	n^{-1}	$n^{\frac{-1}{1-2\theta}}$	linear

- We have a spectra covering "almost" all convex functions in finite dimensions[1] .
- Up to now, the infinite dimensional setting is less understood.

[1] Bolte, Daniilidis, Ley, Mazet - 2010

Contents

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Theorem: Convergence for Landweber's algorithm

The sequence generated by $x_{n+1} = x_n - \tau A^*(Ax_n - y)$ strongly converges to $x_0^\dagger$, the projection of x_0 onto the minimizers of f.
The rates for $\|x_n - x_0^\dagger\|$ depend uniquely on the *regularity* of $x_0 - x_0^\dagger$.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Theorem: Convergence for Landweber's algorithm

The sequence generated by $x_{n+1} = x_n - \tau A^*(Ax_n - y)$ strongly converges to $x_0^\dagger$, the projection of x_0 onto the minimizers of f.
The rates for $\|x_n - x_0^\dagger\|$ depend uniquely on the *regularity* of $x_0 - x_0^\dagger$.

Definition: Regularity space

For $\mu > 0$, define the regularity space $X_\mu := R((A^*A)^\mu)$. It is a normed space, equipped with $\|x\|_{X_\mu} = \|(A^*A)^{-\mu}x\|$.

It is a decreasing family of dense subspaces of $\operatorname{Ker} A^\perp$.

Example: Sobolev regularity

If $X = Y = L^2([0, 2\pi])$ and A is the integration operator, then $X_\mu = H^{2\mu}([0, 2\pi])$.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Theorem: Convergence for Landweber's algorithm

The sequence generated by $x_{n+1} = x_n - \tau A^*(Ax_n - y)$ strongly converges to $x_0^\dagger$, the projection of x_0 onto the minimizers of f.
If $x_0 - x_0^\dagger \in X_\mu$, then the convergence is polynomial:

- $\|x_n - x_0^\dagger\| = O\left(n^{-\mu}\right)$
- $f(x_n) - \inf f = O\left(n^{-(1+2\mu)}\right)$

$x_0 - x_0^\dagger \in X_\mu$ is called a source or regularity condition.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^{\perp}$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Theorem (Haraux, Jendoubi - 2012)

Answer: No.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Theorem (Haraux, Jendoubi - 2012)

The function f satisfies a Lojasiewicz inequality iff $R(A)$ is closed.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^\perp$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Theorem (Haraux, Jendoubi - 2012)

The function f satisfies a Lojasiewicz inequality iff $R(A)$ is closed.

Remark: The rates depend on whether $x_0 \in x_0^\dagger + X_\mu$. If so, the whole sequence remains there.

The Lojasiewicz inequality

Definition (Lojasiewicz, 1965)

Let $x^* \in \operatorname{argmin} f$ and $\theta \in]0, 1/2]$. We say that f has the Lojasiewicz property at x^* with exponent θ if $\exists C > 0$, $\delta > 0$ s.t.

$$x \in x^* + \delta\mathbb{B} \text{ and } f(x) > \inf f \Rightarrow (f(x) - \inf f)^{1-\theta} \leq C\|\partial f(x)\|_{\_}$$

In our convex case, equivalent to an error bound[1]

$$x \in x^* + \delta\mathbb{B} \text{ and } f(x) > \inf f \Rightarrow C' \operatorname{dist}(x, \operatorname{argmin} f)^{1/\theta} \leq f(x) - \inf f$$

The Lojasiewicz exponent θ governs the local geometry of f, and then the rates of convergence.

[1] Bolte, Nguyen, Peypouquet, Suter - 2015

The Lojasiewicz inequality

Definition (Lojasiewicz, 1965)

Let $x^* \in \operatorname{argmin} f$ and $\theta \in]0, 1/2]$. We say that f has the Lojasiewicz property at x^* with exponent θ if $\exists C > 0$, $\delta > 0$ s.t.

$$x \in x^* + \delta\mathbb{B} \text{ and } f(x) > \inf f \Rightarrow (f(x) - \inf f)^{1-\theta} \leq C\|\partial f(x)\|_{-}$$

In our convex case, equivalent to an error bound[1]

$$x \in x^* + \delta\mathbb{B} \text{ and } f(x) > \inf f \Rightarrow C' \operatorname{dist}(x, \operatorname{argmin} f)^{1/\theta} \leq f(x) - \inf f$$

The Lojasiewicz exponent θ governs the local geometry of f, and then the rates of convergence.

[1] Bolte, Nguyen, Peypouquet, Suter - 2015

Definition (Lojasiewicz, 1965)

Let $x^* \in \operatorname{argmin} f$ and $\theta \in]0, 1/2]$. We say that f has the Lojasiewicz property at x^* with exponent θ if $\exists C > 0$, ~~$\delta > 0$~~ $\Omega \subset X$ s.t.

$$x \in x^* + \Omega \text{ and } f(x) > \inf f \Rightarrow (f(x) - \inf f)^{1-\theta} \leq C \|\partial f(x)\|_-$$

In our convex case, equivalent to an error bound[1]

$$x \in x^* + \Omega \text{ and } f(x) > \inf f \Rightarrow C' \operatorname{dist}(x, \operatorname{argmin} f)^{1/\theta} \leq f(x) - \inf f$$

The Lojasiewicz exponent θ governs the local geometry of f, and then the rates of convergence.

[1] Bolte, Nguyen, Peypouquet, Suter - 2015

The Lojasiewicz inequality: Convergence result

Theorem

Let $(x_n)_{n\in\mathbb{N}}$ be generated by the Forward-Backward method, and assume that $(x_n)_{n\in\mathbb{N}} \subset x^* + \Omega$ where the Lojasiewicz inequality holds. Then x_n converges strongly to a minimizer $x^\dagger$ of f. Moreover, depending on the Lojasiewicz exponent θ:

1. If $\theta = 1/2$, then both iterates $\|x_n - x^\dagger\|$ and values $f(x_n) - \inf f$ converge linearly.
2. If $\theta \in]0, 1/2[$, then iterates and values converge polynomially:
   - $\|x_n - x^\dagger\| = O\left(n^{\frac{-\theta}{1-2\theta}}\right)$
   - $f(x_n) - \inf f = O\left(n^{\frac{-1}{1-2\theta}}\right)$

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^{\perp}$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^{\perp}$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Theorem (Garrigos - 2016)

Answer: Yes.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^{\perp}$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Theorem (Garrigos - 2016)

On each X_μ, the least squares f satisfies the Lojasiewicz inequality with the exponent $\theta = \frac{\mu}{1+2\mu}$.

Least squares: $f(x) = \frac{1}{2}\|Ax - y\|^2$

Assume that $R(A)$ is not closed and $y \in R(A) + R(A)^{\perp}$.

Question: Can the polynomial rates for Landweber's algorithm be explained by some Lojasiewicz inequality on $f(x) = \frac{1}{2}\|Ax - y\|^2$?

Theorem (Garrigos - 2016)

On each X_μ, the least squares f satisfies the Lojasiewicz inequality with the exponent $\theta = \frac{\mu}{1+2\mu}$.

This Lojasiewicz exponent is **tight**, and gives exactly the rates of convergence we expect for Landweber's method.

function	convex	$\theta \in]0, \frac{1}{2}[$	$\theta = \frac{1}{2}$
iterates	weak cv	$n^{\frac{-\theta}{1-2\theta}}$	linear
values	n^{-1}	$n^{\frac{-1}{1-2\theta}}$	linear

Conclusion

If you had to remember ONE thing

You have a descent (dissipative?) algorithm?
Strong convexity gives you strong convergence and linear rates?

Try to use the Lojasiewicz inequality with $\theta = 1/2$:

$$C \operatorname{dist}(x, \operatorname{argmin} f)^2 \leq f(x) - \inf f$$

- The Lojasiewicz inequality is not reserved to non-convex optimization.

Discussion

- The Lojasiewicz inequality is not reserved to non-convex optimization.
- It sheds a new light on *a priori* unrelated convergence rates results.

Discussion

- The Lojasiewicz inequality is not reserved to non-convex optimization.
- It sheds a new light on *a priori* unrelated convergence rates results.
- In some cases finding θ and C is not an easy task (LASSO). A general *quantitative* structural result would be welcomed.

Discussion

- The Lojasiewicz inequality is not reserved to non-convex optimization.
- It sheds a new light on *a priori* unrelated convergence rates results.
- In some cases finding θ and C is not an easy task (LASSO). A general *quantitative* structural result would be welcomed.
- In infinite dimension, convergence of descent methods for minimizing

$$\alpha\|x\|_p^p + \frac{1}{2}\|Ax - y\|^2, \quad p \in [1,2]$$

generally requires "«source-like»" conditions.

Discussion

- The Lojasiewicz inequality is not reserved to non-convex optimization.
- It sheds a new light on *a priori* unrelated convergence rates results.
- In some cases finding θ and C is not an easy task (LASSO). A general *quantitative* structural result would be welcomed.
- In infinite dimension, convergence of descent methods for minimizing

$$\alpha\|x\|_p^p + \frac{1}{2}\|Ax - y\|^2, \quad p \in [1, 2]$$

  generally requires "«source-like»" conditions.
- The Lojasiewicz inequality has been used for inertial/accelerated methods. What are the rates?

Thanks for your attention !