BISTA: a Bregmanian proximal gradient method without the global Lipschitz continuity assumption

Daniel Reem
(joint work with Simeon Reich and Alvaro De Pierro)

Department of Mathematics, The Technion, Haifa, Israel

E-mail: dream@technion.ac.il

http://w3.impa.br/~dream

Nonlinear Analysis and Optimization Seminar,
The Technion, Haifa, Israel
(50-60 minutes)
Goal: to estimate \( \inf \{ F(x) : x \in C \} \)
where \( F = f + g \)
\( C \) is a closed and convex subset of some space \( X \), say \( \mathbb{R}^n \),
f is smooth, g possibly not, both are convex.
Minimization of a separable function

**Goal:** to estimate
Goal: to estimate

$$\inf\{F(x) : x \in C\}$$
Goal: to estimate

\[ \inf \{ F(x) : x \in C \} \]

where
Goal: to estimate

\[ \inf \{ F(x) : x \in C \} \]

where

- \( F = f + g \)
Minimization of a separable function

Goal: to estimate

\[ \inf \{ F(x) : x \in C \} \]

where

- \( F = f + g \)
- \( C \) is a closed and convex subset of some space \( X \), say \( \mathbb{R}^n \)
Goal: to estimate

\[ \inf \{ F(x) : x \in C \} \]

where

- \( F = f + g \)
- \( C \) is a closed and convex subset of some space \( X \), say \( \mathbb{R}^n \)
- \( f \) is smooth, \( g \) possibly not, both are convex.
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:

- inverse problems
- image processing
- compressed sensing
- machine learning
- more
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:

- **inverse problems**
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:

- **inverse problems**
- **image processing**
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:

- **inverse problems**
- **image processing**
- **compressed sensing**
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:

- **inverse problems**
- **image processing**
- **compressed sensing**
- **machine learning**
Motivation

Such a minimization problem appears in the solution process of many theoretical and practical problems, including:

- inverse problems
- image processing
- compressed sensing
- machine learning
- more
A typical scenario: $\ell_1$ regularization in signal processing

Here $\inf_{x \in \mathbb{R}^n} (\|Ax - b\|_2 + \lambda \|x\|_1)$

Now $g$ is not smooth.

$A$ is an $m \times n$ matrix. $m$ and $n$ are large positive integers.

$b \in \mathbb{R}^m$ is known.

$\lambda$ is a regularization parameter.
A typical scenario: $\ell_1$ regularization in signal processing

- Here
A typical scenario: $\ell_1$ regularization in signal processing

Here

$$\inf_{x \in \mathbb{R}^n} \left( \| Ax - b \|^2 + \lambda \| x \|_1 \right)$$

Now $g$ is not smooth. $A$ is an $m \times n$ matrix. $m$ and $n$ are large positive integers. $b \in \mathbb{R}^m$ is known. $\lambda$ is a regularization parameter.
A typical scenario: $\ell_1$ regularization in signal processing

- Here

$$\inf_{x \in \mathbb{R}^n} (\|Ax - b\|^2 + \lambda \|x\|_1)$$

- Now $g$ is not smooth.
A typical scenario: $\ell_1$ regularization in signal processing

- Here

$$\inf_{x \in \mathbb{R}^n} (\|Ax - b\|^2 + \lambda \|x\|_1)$$

- Now $g$ is not smooth.

- $A$ is an $m \times n$ matrix.
A typical scenario: $\ell_1$ regularization in signal processing

- Here

$$
\inf_{x \in \mathbb{R}^n} \left( \|Ax - b\|^2 + \lambda \|x\|_1 \right)
$$

- Now $g$ is not smooth.

- $A$ is an $m \times n$ matrix.

- $m$ and $n$ are large positive integers.
A typical scenario: $\ell_1$ regularization in signal processing

- Here

$$\inf_{x \in \mathbb{R}^n} (\|Ax - b\|^2 + \lambda \|x\|_1)$$

- Now $g$ is not smooth.

- $A$ is an $m \times n$ matrix.

- $m$ and $n$ are large positive integers.

- $b \in \mathbb{R}^m$ is known.
A typical scenario: $\ell_1$ regularization in signal processing

- Here

$$\inf_{x \in \mathbb{R}^n} (\|Ax - b\|^2 + \lambda \|x\|_1)$$

- Now $g$ is not smooth.

- $A$ is an $m \times n$ matrix.

- $m$ and $n$ are large positive integers.

- $b \in \mathbb{R}^m$ is known.

- $\lambda$ is a regularization parameter
The proximal gradient method

Among the methods used for solving the minimization problem.

Basic form:

\[ x_{k} := \text{argmin}_{x \in C} (F(x) + c_{k} \| x - x_{k-1} \|^{2}) \], \quad k \geq 2

A more general form:

\[ x_{k} := \text{argmin}_{x \in C} (F_{k}(x) + c_{k} B(x, x_{k-1})) \], \quad k \geq 2

Here \( F_{k} \) is an approximation to \( F = f + g \), e.g.,

\[ F_{k}(x) := f(x_{k-1}) + \langle f'(x_{k-1}), x - x_{k-1} \rangle + g(x) \].

\( B \) is a Bregman divergence induced by some function \( b \) (more on that: later).
The proximal gradient method

- **Among the methods used** for solving the minimization problem.
Among the methods used for solving the minimization problem.

Basic form:

\[
\begin{align*}
\textstyle
x_{k+1} &= \argmin_{x \in C} (F(x) + c_k \|x - x_k - 1\|_2), \quad k \geq 2 \\
\textstyle
x_k &= \argmin_{x \in C} (F_k(x) + c_k B(x, x_k - 1)), \quad k \geq 2.
\end{align*}
\]
The proximal gradient method

- **Among the methods used** for solving the minimization problem.
- Basic form:
  - $x_1$ is given

$$x_k := \arg\min_{x \in C} \left( F(x) + c_k \| x - x_{k-1} \|_2 \right), \quad k \geq 2$$

A more general form:

$$x_k := \arg\min_{x \in C} \left( F_k(x) + c_k B(x, x_{k-1}) \right), \quad k \geq 2.$$
The proximal gradient method

- **Among the methods used** for solving the minimization problem.
- Basic form:
  - $x_1$ is given
  $$x_k := \arg\min_{x \in C} (F(x) + c_k \|x - x_{k-1}\|^2), \quad k \geq 2$$
The proximal gradient method

- **Among the methods used** for solving the minimization problem.

  - Basic form:
    - $x_1$ is given
    - 
      $$x_k := \arg\min_{x \in \mathcal{C}} (F(x) + c_k \|x - x_{k-1}\|^2), \quad k \geq 2$$

  - A more general form:
    - 
      $$x_k := \arg\min_{x \in \mathcal{C}} (F_k(x) + c_k B(x, x_{k-1})), \quad k \geq 2.$$
The proximal gradient method

- **Among the methods used** for solving the minimization problem.

- Basic form:
  - $x_1$ is given
  
  $$x_k := \arg\min_{x \in C} (F(x) + c_k \|x - x_{k-1}\|^2), \quad k \geq 2$$

- A more general form:
  
  $$x_k := \arg\min_{x \in C} (F_k(x) + c_k B(x, x_{k-1})), \quad k \geq 2.$$  

- Here $F_k$ is an approximation to $F = f + g$, e.g.,

  $$F_k(x) := f(x_{k-1}) + \langle f'(x_{k-1}), x - x_{k-1} \rangle + g(x).$$
The proximal gradient method

• **Among the methods used** for solving the minimization problem.

• Basic form:
  • $x_1$ is given
  
  $$x_k := \arg\min_{x \in C} (F(x) + c_k \|x - x_{k-1}\|^2), \quad k \geq 2$$

• A more general form:
  
  $$x_k := \arg\min_{x \in C} (F_k(x) + c_k B(x, x_{k-1})), \quad k \geq 2.$$ 

• Here $F_k$ is an approximation to $F = f + g$, e.g.,
  
  $$F_k(x) := f(x_{k-1}) + \langle f'(x_{k-1}), x - x_{k-1} \rangle + g(x).$$

• $B$ is a Bregman divergence induced by some function $b$ (**more on that: later**).
Proximal algorithms: a very common assumption

The assumption does not always hold, thus casting a limitation on many prox algorithms.
Proximal algorithms: a very common assumption

- $f'$ is globally Lipschitz continuous on $C$ (or on $X$), namely: there exists $L(f') \geq 0$ such that

\[
\|f'(x) - f'(y)\| \leq L(f') \|x - y\|, \quad \forall x, y \in C
\]
Proximal algorithms: a very common assumption

- $f'$ is globally Lipschitz continuous on $C$ (or on $X$), namely: there exists $L(f') \geq 0$ such that

\[ \|f'(x) - f'(y)\| \leq L(f')\|x - y\|, \quad \forall x, y \in C \]
Proximal algorithms: a very common assumption

- **$f'$ is globally Lipschitz continuous on $C$ (or on $X$),** namely: there exists $L(f') \geq 0$ such that

  $$\|f'(x) - f'(y)\| \leq L(f')\|x - y\|, \quad \forall x, y \in C$$

- **This assumption does not always hold,** thus casting a limitation on many prox algorithms
Our results: schematic description


Prox operator is induced by a Bregman divergence, a core new contribution.

We decompose $\mathbb{C} = \bigcup_{k=1}^{\infty} S_k$.

Proximal iteration: on $S_k$ and not globally (on the whole $C$ or $X$); hence more flexibility, better chance that good properties are satisfied (e.g., frequently all the $S_k$ are bounded).
Introducing **BISTA**:

- **BISTA** introduces a new proximal gradient method.
- The prox operator is induced by a Bregman divergence.
- A core new contribution:
  - We decompose $C = \bigcup_{k=1}^{\infty} S_k$.
  - Proximal iteration: on $S_k$ and not globally (on the whole $C$ or $X$); hence more flexibility, better chance that good properties are satisfied (e.g., frequently all the $S_k$ are bounded).
Introducing **BISTA**: a new proximal gradient method
Our results: schematic description

- Introducing **BISTA**: a new proximal gradient method
- Prox operator is induced by a **Bregman divergence**
Our results: schematic description

- Introducing **BISTA**: a new proximal gradient method
- Prox operator is induced by a **Bregman divergence**
- A core new contribution:
Introducing **BISTA**: a new proximal gradient method

Prox operator is induced by a **Bregman divergence**

**a core new contribution:**

- We decompose $C = \bigcup_{k=1}^{\infty} S_k$
Introducing **BISTA**: a new proximal gradient method

Prox operator is induced by a **Bregman divergence**

**a core new contribution:**

- We decompose \( C = \bigcup_{k=1}^{\infty} S_k \)
- Proximal iteration: on \( S_k \) **and not globally** (on the whole \( C \) or \( X \))
Our results: schematic description

- Introducing **BISTA**: a new proximal gradient method
- Prox operator is induced by a **Bregman divergence**
- A core new contribution:
  - We decompose $C = \bigcup_{k=1}^{\infty} S_k$
  - Proximal iteration: on $S_k$ and not globally (on the whole $C$ or $X$);
  - Hence more flexibility, better chance that good properties are satisfied (e.g., frequently all the $S_k$ are bounded)
The origin of BISTA

ISTA (also known as ISTA): a method suggested by Beck-Teboulle (2009).

BISTA significantly extends ISTA.

BISTA = Bregmanian ISTA [or Brazilian ISTA :-)]

Preliminary version:
in 2014-2015, during the postdoc in Brazil (I worked with Alvaro De Pierro)
The origin of BISTA

- ISTA:

BISTA significantly extends ISTA. ISTA = Bregmanian ISTA [or Brazilian ISTA :-)]

Preliminary version: in 2014-2015, during the postdoc in Brazil (I worked with Alvaro De Pierro)
The origin of BISTA

- **ISTA**: a method suggested by Beck-Teboulle (2009)
The origin of BISTA

- **ISTA**: a method suggested by Beck-Teboulle (2009)
- **BISTA**: significantly extends ISTA
The origin of BISTA

- **ISTA**: a method suggested by Beck-Teboulle (2009)
- **BISTA** significantly extends **ISTA**
- **BISTA** = Bregmanian ISTA [or Brazilian ISTA :-)]
The origin of BISTA

- **ISTA**: a method suggested by Beck-Teboulle (2009)

- **BISTA** significantly extends ISTA

- **BISTA** = Bregmanian ISTA [or Brazilian ISTA :-) ]

- **Preliminary version**: in 2014-2015, during the postdoc in Brazil (I worked with Alvaro De Pierro)
Our results: schematic description (Cont.)

Relatively general setting: real reflexive Banach space $X$ (Hilbertian in practice).

Constrained minimization: $C \subseteq X$

$f'$ is not necessary globally Lipschitz continuous (should be Lipschitz continuous only on a subset of $S_k$).

Several convergence results (under some assumptions), for example:

- Non-asymptotic (in the function values) convergence rate of $O\left(\frac{1}{k}\right)$
- Or a rate arbitrarily close to $O\left(\frac{1}{k}\right)$,

Weak convergence.
Relatively general setting: real reflexive Banach space $X$ (Hilbertian in practice)
Our results: schematic description (Cont.)

- **Relatively general setting:** real reflexive Banach space $X$ (Hilbertian in practice)

- **Constrained minimization:** $C \subseteq X$
- **Relatively general setting:** real reflexive Banach space $X$ (Hilbertian in practice)

- **Constrained minimization:** $C \subseteq X$

- $f'$ is not necessary globally Lipschitz continuous (should be Lipschitz continuous only on a subset of $S_k$)
Our results: schematic description (Cont.)

- **Relatively general setting**: real reflexive Banach space $X$ (Hilbertian in practice)

- **Constrained minimization**: $C \subseteq X$

- $f'$ is not necessary globally Lipschitz continuous (should be Lipschitz continuous only on a subset of $S_k$)

- **Several convergence results** (under some assumptions), for example:
Relatively general setting: real reflexive Banach space $X$ (Hilbertian in practice)

Constrained minimization: $C \subseteq X$

$f'$ is not necessary globally Lipschitz continuous (should be Lipschitz continuous only on a subset of $S_k$)

Several convergence results (under some assumptions), for example:

- Non-asymptotic (in the function values) convergence rate of $O(1/k)$ or a rate arbitrarily close to $O(1/k)$,
Our results: schematic description (Cont.)

- **Relatively general setting:** real reflexive Banach space $X$ (Hilbertian in practice)

- **Constrained minimization:** $C \subseteq X$

- $f'$ is not necessary globally Lipschitz continuous (should be Lipschitz continuous only on a subset of $S_k$)

- **Several convergence results** (under some assumptions), for example:
  - Non-asymptotic **(in the function values)** convergence rate of $O(1/k)$ or a rate arbitrarily close to $O(1/k)$,
  - Weak convergence
Our results: schematic description (Cont.)

Some auxiliary results of independent interest:

- Generalization of a key lemma in Beck-Teboulle 2009
- Sufficient conditions for the minimizer which appears in the proximal operation to be an interior point

A general and useful stability principle: given a uniformly continuous real function defined on arbitrary metric space, if we slightly change the objective set over which the optimal (extreme) values of the function are computed, then these values vary slightly.

This stability principle suggests a general scheme for tackling a wide class of non-convex and non-smooth optimization problems.
Some auxiliary results \textit{of independent interest}:
Some auxiliary results of independent interest:

- Generalization of a key lemma in Beck-Teboulle 2009
Some auxiliary results of independent interest:

- Generalization of a key lemma in Beck-Teboulle 2009
- Sufficient conditions for the minimizer which appears in the proximal operation to be an interior point
Some auxiliary results **of independent interest:**

- Generalization of a key lemma in Beck-Teboulle 2009
- Sufficient conditions for the minimizer which appears in the proximal operation to be an interior point
- **A general and useful stability principle:**
Some auxiliary results of independent interest:

- Generalization of a key lemma in Beck-Teboulle 2009
- Sufficient conditions for the minimizer which appears in the proximal operation to be an interior point
- A general and useful stability principle: given a uniformly continuous real function defined on arbitrary metric space,
Some auxiliary results of independent interest:

- Generalization of a key lemma in Beck-Teboulle 2009
- Sufficient conditions for the minimizer which appears in the proximal operation to be an interior point
- A general and useful stability principle: given a uniformly continuous real function defined on arbitrary metric space, if we slightly change the objective set over which the optimal (extreme) values of the function are computed, then these values vary slightly.
Some auxiliary results of independent interest:

- Generalization of a key lemma in Beck-Teboulle 2009
- Sufficient conditions for the minimizer which appears in the proximal operation to be an interior point

A general and useful stability principle: given a uniformly continuous real function defined on arbitrary metric space, if we slightly change the objective set over which the optimal (extreme) values of the function are computed, then these values vary slightly.

This stability principle suggests a general scheme for tackling a wide class of non-convex and non-smooth optimization problems.
Bregman divergences (Bregman distances)

Let \( b : X \to (-\infty, \infty] \)

Assume that:

\[ U := \text{Int}(\text{dom}(b)) \neq \emptyset \]

where \( \text{dom}(b) := \{ x \in X : b(x) < \infty \} \).

\( b \) is Gâteaux differentiable in \( U \).

\( b \) is convex and lower semicontinuous on \( X \) and strictly convex on \( \text{dom}(b) \).

We refer to \( b \) as a semi-Bregman function.

\( B \) is the (semi-)Bregman divergence associated with \( b \):

\[
B(x, y) := \begin{cases}
    b(x) - b(y) - \langle b'(y), x - y \rangle, & (x, y) \in \text{dom}(b) \times U, \\
    \infty & \text{otherwise}.
\end{cases}
\]

Here \( b'(y) \in X^* \) and \( \langle b'(y), x - y \rangle := b'(y)(x - y) \).
Bregman divergences (Bregman distances)

- Let $b : X \rightarrow (\mathbb{R}, \infty]$
Let $b : X \to (-\infty, \infty]$

Assume that:

- $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{x \in X : b(x) < \infty\}$
- $b$ is Gâteaux differentiable in $U$
- $b$ is convex and lower semicontinuous on $X$ and strictly convex on $\text{dom}(b)$.

We refer to $b$ as a semi-Bregman function.

$B$ is the (semi-)Bregman divergence associated with $b$:

$$B(x, y) := \begin{cases} b(x) - b(y) - \langle b'(y), x - y \rangle, & (x, y) \in \text{dom}(b) \times U, \\ \infty, & \text{otherwise} \end{cases}$$

Here $b'(y) \in X^*$ and $\langle b'(y), x - y \rangle := b'(y)(x - y)$. 

BISTA 25 March 2018 11 / 29
Let $b : X \rightarrow (-\infty, \infty]$

Assume that:
- $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{x \in X : b(x) < \infty\}$
Let $b : X \to (-\infty, \infty]$.

Assume that:

- $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{x \in X : b(x) < \infty\}$
- $b$ is Gâteaux differentiable in $U$. 

The Bregman divergence associated with $b$ is denoted as $B(x, y)$.
Bregman divergences (Bregman distances)

Let $b : X \rightarrow (-\infty, \infty]$

Assume that:

- $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{ x \in X : b(x) < \infty \}$
- $b$ is Gâteaux differentiable in $U$.
- $b$ is convex and lower semicontinuous on $X$ and strictly convex on $\text{dom}(b)$. 

$B$ is the (semi-)Bregman divergence associated with $b$:

$$B(x, y) := \begin{cases} b(x) - b(y) - \langle b'(y), x - y \rangle, & (x, y) \in \text{dom}(b) \times U \\ \infty, & \text{otherwise} \end{cases}$$

Here $b'(y) \in X^*$ and $\langle b'(y), x - y \rangle := b'(y)(x - y)$.
Bregman divergences (Bregman distances)

- Let $b : X \rightarrow (-\infty, \infty]$
- Assume that:
  - $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{x \in X : b(x) < \infty\}$
  - $b$ is Gâteaux differentiable in $U$.
  - $b$ is convex and lower semicontinuous on $X$ and strictly convex on $\text{dom}(b)$.
- We refer to $b$ as a semi-Bregman function
Let $b : X \to (-\infty, \infty]$ 

Assume that:
- $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{x \in X : b(x) < \infty\}$
- $b$ is Gâteaux differentiable in $U$.
- $b$ is convex and lower semicontinuous on $X$ and strictly convex on $\text{dom}(b)$.

We refer to $b$ as a **semi-Bregman function**

$B$ is the **(semi-)Bregman divergence** associated with $b$: 

$B(x, y) := \begin{cases} b(x) - b(y) - \langle b'(y), x - y \rangle, & (x, y) \in \text{dom}(b) \times U, \\ \infty & \text{otherwise} \end{cases}$
Let $b : X \to (-\infty, \infty]$.

Assume that:
- $U := \text{Int}(\text{dom}(b)) \neq \emptyset$, where $\text{dom}(b) := \{x \in X : b(x) < \infty\}$
- $b$ is Gâteaux differentiable in $U$.
- $b$ is convex and lower semicontinuous on $X$ and strictly convex on $\text{dom}(b)$.

We refer to $b$ as a **semi-Bregman function**.

$B$ is the **(semi-)Bregman divergence** associated with $b$:

$$B(x, y) := \begin{cases} b(x) - b(y) - \langle b'(y), x - y \rangle, & (x, y) \in \text{dom}(b) \times U, \\ \infty, & \text{otherwise.} \end{cases}$$
Bregman divergences (Bregman distances)

- Let \( b : X \rightarrow (-\infty, \infty] \)

- Assume that:
  - \( U := \mathrm{Int}(\text{dom}(b)) \neq \emptyset \), where \( \text{dom}(b) := \{ x \in X : b(x) < \infty \} \)
  - \( b \) is Gâteaux differentiable in \( U \).
  - \( b \) is convex and lower semicontinuous on \( X \) and strictly convex on \( \text{dom}(b) \).

- We refer to \( b \) as a **semi-Bregman function**

- \( B \) is the **(semi-)Bregman divergence** associated with \( b \):

\[
B(x, y) := \begin{cases} 
    b(x) - b(y) - \langle b'(y), x - y \rangle, & (x, y) \in \text{dom}(b) \times U, \\
    \infty, & \text{otherwise}.
\end{cases}
\]

Here \( b'(y) \in X^* \) and \( \langle b'(y), x - y \rangle := b'(y)(x - y) \)
Suppose that:

- $C$ is a convex subset
- $\emptyset \neq S \subseteq C$ (S is not necessarily convex)
- $b : C \rightarrow \mathbb{R}$ is called strongly convex on $S$ if there exists $\mu > 0$ such that for each $(x, y) \in S^2$ and each $\lambda \in (0, 1)$
  $$b(\lambda x + (1 - \lambda) y) \leq \lambda b(x) + (1 - \lambda) b(y) - \frac{1}{2} \mu \lambda (1 - \lambda) \|x - y\|^2.$$ 

In other words, $b$ satisfies a stronger condition than convexity.

Many well-known Bregman functions are strongly convex on bounded subsets of their effective domain, e.g., the negative Boltzmann-Gibbs-Shannon entropy

$$b(x) = \sum_{k=1}^{n} x_k \log(x_k).$$
Strongly convex functions: reminder

Suppose that:

- Suppose that:
Suppose that:

- $C$ is a convex subset

Strongly convex functions: reminder
Strongly convex functions: reminder

- Suppose that:
  - $C$ is a convex subset
  - $\emptyset \neq S \subseteq C$ ($S$ is not necessarily convex)
Suppose that:
- $C$ is a convex subset
- $\emptyset \neq S \subseteq C$ ($S$ is not necessarily convex)

$b : C \rightarrow \mathbb{R}$ is called strongly convex on $S$ if there exists $\mu > 0$ such that for each $(x, y) \in S^2$ and each $\lambda \in (0, 1)$
Strongly convex functions: reminder

Suppose that:
- $C$ is a convex subset
- $\emptyset \neq S \subseteq C$ ($S$ is not necessarily convex)

$b : C \to \mathbb{R}$ is called **strongly convex on** $S$ if there exists $\mu > 0$ such that for each $(x, y) \in S^2$ and each $\lambda \in (0, 1)$

$$b(\lambda x + (1 - \lambda)y) \leq \lambda b(x) + (1 - \lambda)b(y) - 0.5\mu \lambda(1 - \lambda)\|x - y\|^2.$$
Suppose that:

- C is a convex subset
- ∅ ≠ S ⊆ C (S is not necessarily convex)

\( b : C \rightarrow \mathbb{R} \) is called **strongly convex on** \( S \) if there exists \( \mu > 0 \) such that for each \( (x, y) \in S^2 \) and each \( \lambda \in (0, 1) \)

\[
 b(\lambda x + (1 - \lambda)y) \leq \lambda b(x) + (1 - \lambda)b(y) - 0.5\mu\lambda(1 - \lambda)\|x - y\|^2.
\]

In other words, \( b \) **satisfies a stronger condition than convexity**
Suppose that:
- $C$ is a convex subset
- $\emptyset \neq S \subseteq C$ ($S$ is not necessarily convex)

$b : C \to \mathbb{R}$ is called **strongly convex on $S$** if there exists $\mu > 0$ such that for each $(x, y) \in S^2$ and each $\lambda \in (0, 1)$

$$b(\lambda x + (1 - \lambda)y) \leq \lambda b(x) + (1 - \lambda)b(y) - 0.5\mu \lambda (1 - \lambda)\|x - y\|^2.$$ 

In other words, $b$ satisfies a stronger condition than convexity

**Many well-known Bregman functions** are strongly convex on bounded subsets of their effective domain,
Strongly convex functions: reminder

- Suppose that:
  - $C$ is a convex subset
  - $\emptyset \neq S \subseteq C$ ($S$ is not necessarily convex)

- $b : C \to \mathbb{R}$ is called **strongly convex on $S$** if there exists $\mu > 0$ such that for each $(x, y) \in S^2$ and each $\lambda \in (0, 1)$

$$b(\lambda x + (1 - \lambda)y) \leq \lambda b(x) + (1 - \lambda)b(y) - 0.5\mu\lambda(1 - \lambda)\|x - y\|^2.$$

- In other words, $b$ satisfies a stronger condition than convexity

- **Many well-known Bregman functions** are strongly convex on bounded subsets of their effective domain, e.g., the negative **Boltzmann-Gibbs-Shannon entropy** $b(x) = \sum_{k=1}^{n} x_k \log(x_k)$
A few assumptions

\[ C \subseteq \text{dom}(b) \]

\[ f : \text{dom}(b) \rightarrow \mathbb{R} : \text{convex on } \text{dom}(b) , \text{Gateaux differentiable in } U ; \]

\[ g : C \rightarrow (-\infty, \infty] : \text{convex, proper, lower semicontinuous} \]

\[ F(x) := \{ f(x) + g(x), x \in C, \infty, x \not\in C \}. \]

\[ \text{OPT}(F) (\text{the optimal set of } F, \text{namely its set of minimizers}) \text{ is nonempty and contained in } U := \text{Int(dom}(b)). \]
A few assumptions

- $C \subseteq \text{dom}(b)$
A few assumptions

- $C \subseteq \text{dom}(b)$
- $f : \text{dom}(b) \to \mathbb{R}$: convex on $\text{dom}(b)$, Gâteaux differentiable in $U$;
A few assumptions

- $C \subseteq \text{dom}(b)$
- $f : \text{dom}(b) \rightarrow \mathbb{R}$: convex on $\text{dom}(b)$, Gâteaux differentiable in $U$;
- $g : C \rightarrow (-\infty, \infty]$: convex, proper, lower semicontinuous
A few assumptions

- $C \subseteq \text{dom}(b)$
- $f : \text{dom}(b) \rightarrow \mathbb{R}$: convex on $\text{dom}(b)$, Gâteaux differentiable in $U$;
- $g : C \rightarrow (-\infty, \infty]$: convex, proper, lower semicontinuous

$$F(x) := \begin{cases} f(x) + g(x), & x \in C, \\ \infty, & x \notin C. \end{cases}$$
A few assumptions

- $C \subseteq \text{dom}(b)$
- $f : \text{dom}(b) \to \mathbb{R}$: convex on $\text{dom}(b)$, Gâteaux differentiable in $U$;
- $g : C \to (-\infty, \infty]$: convex, proper, lower semicontinuous
- $F(x) := \begin{cases} f(x) + g(x), & x \in C, \\ \infty, & x \notin C. \end{cases}$
- $\text{OPT}(F)$ (the optimal set of $F$, namely its set of minimizers) is nonempty and contained in $U := \text{Int}(\text{dom}(b))$
A few assumptions (Cont.)

\[ C = \bigcup_{k=1}^{\infty} S_k \] (a core new contribution)

For each \( k \):

\[ S_k \text{ is closed, convex, } S_k \cap U \neq \emptyset, S_k \subseteq S_{k+1} \]

\( b \) is strongly convex on \( S_k \) with \( \mu_k > 0 \) and \( \mu_k \geq \mu_{k+1} \)

\( g \) is proper on \( S_k \)

\( f' \) is Lipschitz continuous on \( S_k \cap U \) with a constant \( L(f', S_k \cap U) \)

BISTA 25 March 2018 14 / 29
A few assumptions (Cont.)

- \( C = \bigcup_{k=1}^{\infty} S_k \) (a core new contribution)
A few assumptions (Cont.)

- \( C = \bigcup_{k=1}^{\infty} S_k \) (a core new contribution)
- For each \( k \):

\[ b \text{ is strongly convex on } S_k \text{ with } \mu_k > 0 \text{ and } \mu_k \geq \mu_{k+1} \]

\[ g \text{ is proper on } S_k \]

\[ f' \text{ is Lipschitz continuous on } S_k \cap U \text{ with a constant } L(f', S_k \cap U) \]
A few assumptions (Cont.)

- \( C = \bigcup_{k=1}^{\infty} S_k \) (a core new contribution)

- For each \( k \):
  
  - \( S_k \) is closed, convex, \( S_k \cap U \neq \emptyset \), \( S_k \subseteq S_{k+1} \)
A few assumptions (Cont.)

- $C = \bigcup_{k=1}^{\infty} S_k$ (a core new contribution)

- For each $k$:
  - $S_k$ is closed, convex, $S_k \cap U \neq \emptyset$, $S_k \subseteq S_{k+1}$
  - $b$ is strongly convex on $S_k$ with $\mu_k > 0$ and $\mu_k \geq \mu_{k+1}$
A few assumptions (Cont.)

- $C = \bigcup_{k=1}^{\infty} S_k$ (a core new contribution)

For each $k$:

- $S_k$ is closed, convex, $S_k \cap U \neq \emptyset$, $S_k \subseteq S_{k+1}$
- $b$ is strongly convex on $S_k$ with $\mu_k > 0$ and $\mu_k \geq \mu_{k+1}$
- $g$ is proper on $S_k$
A few assumptions (Cont.)

- \( C = \bigcup_{k=1}^{\infty} S_k \) (a core new contribution)
- For each \( k \):
  - \( S_k \) is closed, convex, \( S_k \cap U \neq \emptyset \), \( S_k \subseteq S_{k+1} \)
  - \( b \) is strongly convex on \( S_k \) with \( \mu_k > 0 \) and \( \mu_k \geq \mu_{k+1} \)
  - \( g \) is proper on \( S_k \)
  - \( f' \) is Lipschitz continuous on \( S_k \cap U \) with a constant \( L(f', S_k \cap U) \)
Given $L_k > 0$, let $Q_{L_k,\mu_k}(x, y) := \langle f'(y), x - y \rangle + L_k \mu_k B(x, y) + g(x)$, $x \in S_k$, $y \in U$.

Assumption: $p_{L_k,\mu_k,S_k}(y) \in S_k \cap U$ for each $k$.
Given $L_k > 0$, let
Given $L_k > 0$, let

$$Q_{L_k, \mu_k, S_k}(x, y) := f(y) + \langle f'(y), x - y \rangle + \frac{L_k}{\mu_k} B(x, y) + g(x), \quad x \in S_k, \ y \in U.$$
A few assumptions (Cont.)

- Given $L_k > 0$, let

$$Q_{L_k,\mu_k,S_k}(x, y) := f(y) + \langle f'(y), x-y \rangle + \frac{L_k}{\mu_k} B(x, y) + g(x), \ x \in S_k, \ y \in U.$$ 

and

$$p_{L_k,\mu_k,S_k}(y) := \arg\min\{Q_{L_k,\mu_k,S_k}(x, y) : \ x \in S_k\},$$
Given $L_k > 0$, let

$$Q_{L_k,\mu_k,S_k}(x, y) := f(y) + \langle f'(y), x-y \rangle + \frac{L_k}{\mu_k} B(x, y) + g(x), \ x \in S_k, \ y \in U.$$ 

and

$$p_{L_k,\mu_k,S_k}(y) := \text{argmin}\{Q_{L_k,\mu_k,S_k}(x, y) : \ x \in S_k\},$$

**Lemma:**
Given $L_k > 0$, let

$$Q_{L_k, \mu_k, S_k}(x, y) := f(y) + \langle f'(y), x - y \rangle + \frac{L_k}{\mu_k} B(x, y) + g(x), \ x \in S_k, \ y \in U,$$

and

$$p_{L_k, \mu_k, S_k}(y) := \text{argmin}\{Q_{L_k, \mu_k, S_k}(x, y) : \ x \in S_k\},$$

**Lemma:** $p_{L_k, \mu_k, S_k}(y)$ exists and is unique

**Assumption:**
Given $L_k > 0$, let

$$Q_{L_k, \mu_k, s_k}(x, y) := f(y) + \langle f'(y), x - y \rangle + \frac{L_k}{\mu_k} B(x, y) + g(x), \quad x \in S_k, \ y \in U,$$

and

$$p_{L_k, \mu_k, s_k}(y) := \operatorname{argmin}\{ Q_{L_k, \mu_k, s_k}(x, y) : \ x \in S_k \},$$

**Lemma:** $p_{L_k, \mu_k, s_k}(y)$ exists and is unique

**Assumption:** $p_{L_k, \mu_k, s_k}(y) \in S_k \cap U$ for each $k$
Remark on the assumptions

These assumptions occur frequently in applications. For instance:

$$\text{OPT}(F) \neq \emptyset$$

if $C$ is compact or $F$ is coercive.

If $C \subseteq U$, then $\text{OPT}(F) \subseteq U$ and $p_{L_k, \mu_k, S_k(y)} \in S_k \cap U$; both $\text{OPT}(F) \subseteq U$ and $p_{L_k, \mu_k, S_k(y)} \in S_k \cap U$ can be satisfied even if $C \not\subseteq U$ (details: in the paper).
Remark on the assumptions

- These assumptions occur frequently in applications
Remark on the assumptions

- These assumptions occur frequently in applications

- For instance:
Remark on the assumptions

- These assumptions occur frequently in applications
- For instance:
  - \( \text{OPT}(F) \neq \emptyset \) if \( C \) is compact or \( F \) is coercive

If \( C \subseteq U \), then \( \text{OPT}(F) \subseteq U \) and \( p_{L_k, \mu_k, S_k}(y) \in S_k \cap U \); both \( \text{OPT}(F) \subseteq U \) and \( p_{L_k, \mu_k, S_k}(y) \in S_k \cap U \) can be satisfied even if \( C \not\subseteq U \) (details: in the paper)
Remark on the assumptions

- **These assumptions occur frequently in applications**

- For instance:
  - \( \text{OPT}(F) \neq \emptyset \) if \( C \) is compact or \( F \) is coercive
  - If \( C \subseteq U \), then \( \text{OPT}(F) \subseteq U \) and \( p_{L_k,\mu_k,s_k}(y) \in S_k \cap U \);

  Both \( \text{OPT}(F) \subseteq U \) and \( p_{L_k,\mu_k,s_k}(y) \in S_k \cap U \) can be satisfied even if \( C \notin U \) (details: in the paper)
Remark on the assumptions

These assumptions occur frequently in applications

For instance:

- $\text{OPT}(F) \neq \emptyset$ if $C$ is compact or $F$ is coercive

- If $C \subseteq U$, then $\text{OPT}(F) \subseteq U$ and $p_{L_k,\mu_k,s_k}(y) \in S_k \cap U$;

- Both $\text{OPT}(F) \subseteq U$ and $p_{L_k,\mu_k,s_k}(y) \in S_k \cap U$ can be satisfied even if $C \not\subseteq U$ (details: in the paper)
BISTA (the Lipschitz step size version)

Input:
A positive number $L$ such that $L \geq L(f', S_1 \cap U)$.

Step 1 (Initialization):
Arbitrary $x_1 \in S_1 \cap U$.

Step $k, k \geq 2$:
$L_k$ is arbitrary such that $L_k \geq \max\{L_k - 1, L(f', S_k \cap U)\}$.

Given $\mu_k > 0$ (a parameter of strong convexity of $b$ on $S_k$), let $x_k := p_{L_k, \mu_k, S_k}(x_k - 1)$. 
Input: A positive number $L_1 \geq L(f', S_1 \cap U)$. 
**Input:** A positive number $L_1 \geq L(f', S_1 \cap U)$.

**Step 1 (Initialization):** arbitrary $x_1 \in S_1 \cap U$. 

BISTA (the Lipschitz step size version)
**Input:** A positive number $L_1 \geq L(f', S_1 \cap U)$.

**Step 1 (Initialization):** arbitrary $x_1 \in S_1 \cap U$.

**Step $k$, $k \geq 2$:**
- $L_k$ is arbitrary such that $L_k \geq \max\{L_{k-1}, L(f', S_k \cap U)\}$. 
**Input:** A positive number \( L_1 \geq L(f', S_1 \cap U) \).

**Step 1 (Initialization):** arbitrary \( x_1 \in S_1 \cap U \).

**Step \( k, k \geq 2 \):**
- \( L_k \) is arbitrary such that \( L_k \geq \max\{L_{k-1}, L(f', S_k \cap U)\} \).
- Given \( \mu_k > 0 \) (a parameter of strong convexity of \( b \) on \( S_k \)), let
BISTA (the Lipschitz step size version)

- **Input:** A positive number $L_1 \geq L(f', S_1 \cap U)$.

- **Step 1 (Initialization):** arbitrary $x_1 \in S_1 \cap U$.

- **Step $k$, $k \geq 2$:**
  - $L_k$ is arbitrary such that $L_k \geq \max\{L_{k-1}, L(f', S_k \cap U)\}$.
  - Given $\mu_k > 0$ (a parameter of strong convexity of $b$ on $S_k$), let
    \[
    x_k := p_{L_k, \mu_k, S_k}(x_{k-1})
    \]
A few relevant works

All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works

- Back-Teboulle 2009,

All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on \( S_k \); even in Tseng the \( S_k \) are very complicated and monotone decreasing (in our case: simple, increasing) and \( f' \) is globally Lipschitz continuous.
A few relevant works

- Back-Teboulle 2009, Cohen 1980,
A few relevant works


All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works

- Back-Teboulle 2009,
- Cohen 1980,
- Tseng 2008 (preprint),
- Bello Cruz and Nghia 2016,

All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works


  All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works


All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works


All of them are significantly different from our work, either in the setting, and/or the method, and/or the results. For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works


- All of them are significantly different from our work, either in

  - the setting,
  - the method,
  - the results

For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works


- All of them are significantly different from our work, either in
  - the setting, and/or

- in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
A few relevant works


- All of them are significantly different from our work, either in
  - the setting, and/or
  - the method, and/or
A few relevant works


- All of them are significantly different from our work, either in
  - the setting, and/or
  - the method, and/or
  - the results
A few relevant works


- All of them are significantly different from our work, either in
  - the setting, and/or
  - the method, and/or
  - the results

- For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$;
A few relevant works


- All of them are significantly different from our work, either in
  - the setting, and/or
  - the method, and/or
  - the results

- For example, in all works with exception of Tseng 2008, the iterations do not depend on $S_k$; even in Tseng the $S_k$ are very complicated and monotone decreasing (in our case: simple, increasing) and $f'$ is globally Lipschitz continuous.
One of the convergence results

Theorem
We impose the previous mentioned assumptions,
Let $x_{\text{opt}} \in \text{OPT}(F)$ be fixed.
Then there exists $k_0 \in \mathbb{N}$ such that for each $k \geq k_0$, we have

$$F(x_{k+1}) - F(x_{\text{opt}}) \leq L_{k+1}B(x_{\text{opt}}, x_k)(k + 1 - k_0)\mu_{k+1}.$$ 

Under further assumptions on $B$ (quite mild: see next slides), if $\lim_{k \to \infty} L_k \mu_k = 0$,
then there exists $z_\infty \in \text{OPT}(F)$ such that $z_\infty = \lim_{k \to \infty} x_k$ weakly.
One of the convergence results

Theorem

- We impose the previous mentioned assumptions,

\[ \text{Let } x_{\text{opt}} \in \text{OPT}(F) \text{ be fixed.} \]

Then there exists \( k_0 \in \mathbb{N} \) such that for each \( k \geq k_0 \), we have

\[ F(x_{k+1}) - F(x_{\text{opt}}) \leq L_{k+1} B(x_{\text{opt}}, x_k) \left( k + 1 - k_0 \right) \mu_{k+1}. \]

Under further assumptions on \( B \) (quite mild: see next slides), if

\[ \lim_{k \to \infty} L_k \mu_k = 0, \]

then there exists \( z_\infty \in \text{OPT}(F) \) such that

\[ z_\infty = \lim_{k \to \infty} x_k \text{ weakly.} \]
One of the convergence results

**Theorem**

- We impose the previous mentioned assumptions,
- Let $x_{\text{opt}} \in \text{OPT}(F)$ be fixed
One of the convergence results

Theorem

- We impose the previous mentioned assumptions,
- Let $x_{opt} \in \text{OPT}(F)$ be fixed
- Then there exists $k_0 \in \mathbb{N}$ such that for each $k \geq k_0$, we have

$$F(x_{k+1}) - F(x_{opt}) \leq L_{k+1} B(x_{opt}, x_k) (k+1 - k_0) \mu_{k+1}.$$
One of the convergence results

**Theorem**

- *We impose the previous mentioned assumptions,*
- *Let* $x_{\text{opt}} \in \text{OPT}(F)$ *be fixed*
- *Then there exists* $k_0 \in \mathbb{N}$ *such that for each* $k \geq k_0$, *we have*

$$F(x_{k+1}) - F(x_{\text{opt}}) \leq \frac{L_{k+1}B(x_{\text{opt}}, x_{k_0})}{(k + 1 - k_0)\mu_{k+1}}.$$
One of the convergence results

**Theorem**

- We impose the previous mentioned assumptions,
- Let $x_{\text{opt}} \in \text{OPT}(F)$ be fixed
- Then there exists $k_0 \in \mathbb{N}$ such that for each $k \geq k_0$, we have
  $$F(x_{k+1}) - F(x_{\text{opt}}) \leq \frac{L_{k+1}B(x_{\text{opt}}, x_{k_0})}{(k + 1 - k_0)\mu_{k+1}}.$$  

- Under further assumptions on $B$ (quite mild: see next slides), if
  $$\lim_{k \to \infty} \frac{L_k}{k\mu_k} = 0,$$
One of the convergence results

Theorem

- We impose the previous mentioned assumptions,
- Let $x_{opt} \in \text{OPT}(F)$ be fixed
- Then there exists $k_0 \in \mathbb{N}$ such that for each $k \geq k_0$, we have
  \[ F(x_{k+1}) - F(x_{opt}) \leq \frac{L_{k+1}B(x_{opt}, x_{k})}{(k + 1 - k_0)\mu_{k+1}}. \]
- Under further assumptions on $B$ (quite mild: see next slides), if
  \[ \lim_{k \to \infty} \frac{L_k}{k\mu_k} = 0, \]
  then there exists $z_\infty \in \text{OPT}(F)$ such that $z_\infty = \lim_{k \to \infty} x_k$ weakly.
Another convergence result

Corollary

The same assumptions as in the previous slide, and also:

\[ f'' \text{ exists, is bounded and uniformly continuous on bounded subsets of } C \cap U, \]

\[ b \text{ is strongly convex on } C \text{ with a strong convexity parameter } \mu > 0. \]

Then we can construct, using BISTA, a sequence \((x_k)_{k=1}^{\infty}\) which converges non-asymptotically to an optimal value, at a rate which can be arbitrary close to \(O(1/k)\).

In particular, for all \(x_{\text{opt}} \in \text{OPT}(F)\), \(q \in (0, 1)\), \(y_0 \in C \cap U\) and \(\alpha > \|f''(y_0)\|\), there are \((x_k)_{k=1}^{\infty}\) and \(k_0 \in \mathbb{N}\) such that

\[ F(x_{k+1}) - F(x_{\text{opt}}) \leq \frac{1 + k}{k+1} - k_0 \cdot \alpha \cdot B(x_{\text{opt}}, x_{k_0}) \cdot \mu (k+1)^{-1} \]

\(\forall k \geq k_0\).

Under further assumptions on \(B\) (quite mild: see next slides), there exists \(z_{\infty} \in \text{OPT}(F)\) such that \(z_{\infty} = \lim_{k \to \infty} x_k\) weakly.
Another convergence result

Corollary

- The same assumptions as in the previous slide, and also:

\[ f'' \] exists, is bounded and uniformly continuous on bounded subsets of \( C \cap U \), \( b \) is strongly convex on \( C \) with a strong convexity parameter \( \mu > 0 \).

Then we can construct, using BISTA, a sequence \( (x_k)_{k=1}^{\infty} \) which converges non-asymptotically to an optimal value, at a rate which can be arbitrarily close to \( O(1/k) \).

In particular, for all \( x_{opt} \in \text{OPT}(F) \), \( q \in (0, 1) \), \( y_0 \in C \cap U \) and \( \alpha > \|f''(y_0)\| \), there are \( (x_k)_{k=1}^{\infty} \) and \( k_0 \in \mathbb{N} \) such that

\[
F(x_{k+1}) - F(x_{opt}) \leq \frac{1}{k+1} - k_0 \cdot \alpha \cdot B(x_{opt}, x_{k_0}) \cdot \frac{1}{\mu (k+1)^{1-q}}, \quad \forall k \geq k_0.
\]

Under further assumptions on \( B \) (quite mild: see next slides), there exists \( z_\infty \in \text{OPT}(F) \) such that \( z_\infty = \lim_{k \to \infty} x_k \) weakly.
Another convergence result

Corollary

- The same assumptions as in the previous slide, and also:
  - \( f'' \) exists, is bounded and uniformly continuous on bounded subsets of \( C \cap U \),
Another convergence result

Corollary

- The same assumptions as in the previous slide, and also:
  - $f''$ exists, is bounded and uniformly continuous on bounded subsets of $C \cap U$,
  - $b$ is strongly convex on $C$ with a strong convexity parameter $\mu > 0$. 

Under further assumptions on $B$ (quite mild: see next slides), there exists $z_\infty \in \text{OPT}(F)$ such that $z_\infty = \lim_{k \to \infty} x_k$ weakly.
Another convergence result

Corollary

- The same assumptions as in the previous slide, and also:
  - \( f'' \) exists, is bounded and uniformly continuous on bounded subsets of \( C \cap U \),
  - \( b \) is strongly convex on \( C \) with a strong convexity parameter \( \mu > 0 \).

- Then we can construct, using BISTA, a sequence \( (x_k)_{k=1}^{\infty} \) which converges non-asymptotically to an optimal value, at a rate which can be arbitrary close to \( O(1/k) \).
Another convergence result

Corollary

- The same assumptions as in the previous slide, and also:
  - $f''$ exists, is bounded and uniformly continuous on bounded subsets of $C \cap U$,
  - $b$ is strongly convex on $C$ with a strong convexity parameter $\mu > 0$.

- Then we can construct, using BISTA, a sequence $(x_k)_{k=1}^\infty$ which converges non-asymptotically to an optimal value, at a rate which can be arbitrary close to $O(1/k)$.

- In particular, for all $x_{opt} \in \text{OPT}(F)$, $q \in (0,1)$, $y_0 \in C \cap U$ and $\alpha > \|f''(y_0)\|$, there are $(x_k)_{k=1}^\infty$ and $k_0 \in \mathbb{N}$ such that

$$F(x_{k+1}) - F(x_{opt}) \leq \frac{k + 1}{k + 1 - k_0} \cdot \frac{\alpha B(x_{opt}, x_{k_0})}{\mu(k + 1)^{1-q}}, \quad \forall k \geq k_0.$$
Another convergence result

Corollary

- The same assumptions as in the previous slide, and also:
  - $f''$ exists, is bounded and uniformly continuous on bounded subsets of $C \cap U$,
  - $b$ is strongly convex on $C$ with a strong convexity parameter $\mu > 0$.

- Then we can construct, using BISTA, a sequence $(x_k)_{k=1}^{\infty}$ which converges non-asymptotically to an optimal value, at a rate which can be arbitrary close to $O(1/k)$.

- In particular, for all $x_{opt} \in \text{OPT}(F)$, $q \in (0, 1)$, $y_0 \in C \cap U$ and $\alpha > \|f''(y_0)\|$, there are $(x_k)_{k=1}^{\infty}$ and $k_0 \in \mathbb{N}$ such that

$$F(x_{k+1}) - F(x_{opt}) \leq \frac{k + 1}{k + 1 - k_0} \cdot \frac{\alpha B(x_{opt}, x_{k_0})}{\mu(k + 1)^{1-q}}, \quad \forall k \geq k_0.$$ 

- Under further assumptions on $B$ (quite mild: see next slides), there exists $z_\infty \in \text{OPT}(F)$ such that $z_\infty = \lim_{k \to \infty} x_k$ weakly.
The further assumptions on $B$ which ensure convergence
The further assumptions on $B$ which ensure convergence

**Assumption**

$B$ has the **limiting difference property**: for each $x \in \text{dom}(b)$ and each sequence $(y_i)_{i=1}^{\infty}$ in $U$, if $(y_i)_{i=1}^{\infty}$ converges weakly to some $y \in U$, then

$$B(x, y) = \lim_{i \to \infty} (B(x, y_i) - B(y, y_i)).$$
The further assumptions on $B$ which ensure convergence

**Assumption**

$B$ has the **limiting difference property**: for each $x \in \text{dom}(b)$ and each sequence $(y_i)_{i=1}^{\infty}$ in $U$, if $(y_i)_{i=1}^{\infty}$ converges weakly to some $y \in U$, then

$$B(x, y) = \lim_{i \to \infty} (B(x, y_i) - B(y, y_i)).$$

**Assumption**

$B$ has **bounded level-sets of the first type**: for each $\gamma \in [0, \infty)$ and each $x \in \text{dom}(b)$, the following set (level-set) is bounded:

$$L_1(x, \gamma) := \{y \in U : B(x, y) \leq \gamma\}.$$
The further assumptions on $B$ (Cont.)

The limiting difference property always holds when $\dim(X) < \infty$; there are infinite-dimensional examples when it holds.

Sufficient conditions for $B$ to have bounded level-sets:

- $b$ is uniformly convex,
- or $b$ satisfies a certain relative uniform convexity assumption with a coercive gauge.
The further assumptions on $B$ (Cont.)

- **The limiting difference property always holds when** $\dim(X) < \infty$; there are infinite-dimensional examples when it holds.
The limiting difference property always holds when \( \dim(X) < \infty \); there are infinite-dimensional examples when it holds.

Sufficient conditions for \( B \) to have bounded level-sets:
The further assumptions on $B$ (Cont.)

- **The limiting difference property always holds when** $\dim(X) < \infty$; there are infinite-dimensional examples when it holds

- **Sufficient conditions for $B$ to have bounded level-sets:**
  - $b$ is uniformly convex, or
The limiting difference property always holds when $\dim(X) < \infty$; there are infinite-dimensional examples when it holds

Sufficient conditions for $B$ to have bounded level-sets:

- $b$ is uniformly convex, or

- $b$ satisfies a certain relative uniform convexity assumption with a coercive gauge
Example: $\ell_p - \ell_1$ minimization

Let: $n \in \mathbb{N}, m \in \mathbb{N}, p \in [2, \infty)$

$X = \mathbb{R}^n, \mathcal{C} = X$ be a linear operator $A: X \rightarrow \mathbb{R}^m$

$z \in \mathbb{R}^m$ given

$\lambda > 0$ is given

For $x \in X$ denote $\|x\|_p := \left( \sum_{k=1}^{n} |x_k|^p \right)^{1/p}$

Goal: to estimate $\inf_{x \in X} \left[ \frac{1}{p} \|Ax - z\|_p \right]$

$\lambda \|x\|_1$
Example: $\ell_p - \ell_1$ minimization

Let:

- $n \in \mathbb{N}$,
- $m \in \mathbb{N}$,
- $p \in [2, \infty)$

$x = \mathbb{R}^n$,

$A: x \to \mathbb{R}^m$ be a linear operator.

$z \in \mathbb{R}^m$ given.

$\lambda > 0$ is given.

For $x \in x$ denote

$\|x\|_p := (\sum_{k=1}^n |x_k|^p)^{1/p}$

Goal: to estimate

$\inf_{x \in X} \left\{ \frac{1}{p} \|Ax - z\|_p + \lambda \|x\|_1 \right\}$
Example: \( \ell_p - \ell_1 \) minimization

Let:

\[ n \in \mathbb{N}, \ m \in \mathbb{N}, \ p \in [2, \infty) \]
Example: $\ell_p-\ell_1$ minimization

Let:

- $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
- $X = \mathbb{R}^n$, $C = X$
Example: $\ell_p$-$\ell_1$ minimization

Let:

- $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
- $X = \mathbb{R}^n$, $C = X$
- $A : X \rightarrow \mathbb{R}^m$ be a linear operator
Example: $\ell_p - \ell_1$ minimization

Let:

- $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
- $X = \mathbb{R}^n$, $C = X$
- $A : X \rightarrow \mathbb{R}^m$ be a linear operator
- $z \in \mathbb{R}^m$ given
Example: $\ell_p - \ell_1$ minimization

Let:

- $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
- $X = \mathbb{R}^n$, $C = X$
- $A : X \rightarrow \mathbb{R}^m$ be a linear operator
- $z \in \mathbb{R}^m$ given
- $\lambda > 0$ is given
Example: $\ell_p$-$\ell_1$ minimization

Let:

- $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
- $X = \mathbb{R}^n$, $C = X$
- $A : X \to \mathbb{R}^m$ be a linear operator
- $z \in \mathbb{R}^m$ given
- $\lambda > 0$ is given

For $x \in X$ denote $\|x\|_p := (\sum_{k=1}^n |x_k|^p)^{1/p}$
Example: $\ell_p - \ell_1$ minimization

- Let:
  - $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
  - $X = \mathbb{R}^n$, $C = X$
  - $A : X \to \mathbb{R}^m$ be a linear operator
  - $z \in \mathbb{R}^m$ given
  - $\lambda > 0$ is given

- For $x \in X$ denote $\|x\|_p := (\sum_{k=1}^{n} |x_k|^p)^{1/p}$

- **Goal**: to estimate
Example: $\ell_p-\ell_1$ minimization

Let:

- $n \in \mathbb{N}$, $m \in \mathbb{N}$, $p \in [2, \infty)$
- $X = \mathbb{R}^n$, $C = X$
- $A : X \to \mathbb{R}^m$ be a linear operator
- $z \in \mathbb{R}^m$ given
- $\lambda > 0$ is given

For $x \in X$ denote $\|x\|_p := (\sum_{k=1}^{n} |x_k|^p)^{1/p}$

Goal: to estimate

$$\inf_{x \in X} \left[ \frac{1}{p} \|Ax - z\|_p^p + \lambda \|x\|_1 \right] ,$$

where $f$ and $g$ are functions.
Example: $\ell_p-\ell_1$ minimization (Cont.)

We use BISTA with $B_{\gamma}(x) := \frac{1}{2} \|x\|_2^2$.

Fix some $\gamma \in (0, 1/(p-2))$ (if $p=2$, then $\gamma$ can be arbitrary positive).

Let $S_k$ be the ball of radius $r_k := k\gamma$ and center 0.

The convergence theorem ensures that $x_k$ converges to an optimal solution.

The non-asymptotic rate of convergence is $O\left(\frac{1}{k^{1-\gamma(p-2)}}\right)$, namely, arbitrarily close to $O(1/k)$. 

BISTA

25 March 2018 24 / 29
Example: \( \ell_p - \ell_1 \) minimization (Cont.)

- \( f' \) is not globally Lipschitz continuous when \( p > 2 \)
Example: $\ell_p - \ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use **BISTA** with $b(x) := \frac{1}{2} \|x\|_2^2$
Example: $\ell_p$-$\ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use BISTA with $b(x) := \frac{1}{2} \|x\|_2^2$
- Fix some $\gamma \in (0, 1/(p - 2))$ (if $p = 2$, then $\gamma$ can be arbitrary positive)
Example: $\ell_p$-$\ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use BISTA with $b(x) := \frac{1}{2} \|x\|_2^2$
- Fix some $\gamma \in (0, 1/(p - 2))$ (if $p = 2$, then $\gamma$ can be arbitrary positive)
- let $S_k$ be the ball of radius $r_k := k^\gamma$ and center 0
Example: $\ell_p$-$\ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use BISTA with $b(x) := \frac{1}{2} \|x\|_2^2$
- Fix some $\gamma \in (0, 1/(p - 2))$ (if $p = 2$, then $\gamma$ can be arbitrary positive)
- let $S_k$ be the ball of radius $r_k := k^\gamma$ and center $0$
- We can take $L_k := (p - 1)2^{p-2} \|A\|^2 k^{\gamma(p-2)}$
Example: $\ell_p - \ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use **BISTA** with $b(x) := \frac{1}{2} \|x\|_2^2$
- Fix some $\gamma \in (0, 1/(p - 2))$ (if $p = 2$, then $\gamma$ can be arbitrary positive)
- Let $S_k$ be the ball of radius $r_k := k^\gamma$ and center 0
- We can take $L_k := (p - 1)2^{p-2}\|A\|_2^2 k^{\gamma(p-2)}$
- The convergence theorem ensures that $x_k$ converges to an optimal solution
Example: $\ell_p$-$\ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use BISTA with $b(x) := \frac{1}{2}\|x\|_2^2$
- Fix some $\gamma \in (0, 1/(p - 2))$ (if $p = 2$, then $\gamma$ can be arbitrary positive)
- Let $S_k$ be the ball of radius $r_k := k^\gamma$ and center 0
- We can take $L_k := (p - 1)2^{p-2}\|A\|_2^2k^\gamma(p-2)$
- The convergence theorem ensures that $x_k$ converges to an optimal solution
- Non-asymptotic rate of convergence is $O\left(\frac{1}{k^{1-\gamma(p-2)}}\right)$,
Example: $\ell_p$-$\ell_1$ minimization (Cont.)

- $f'$ is not globally Lipschitz continuous when $p > 2$
- We use BISTA with $b(x) := \frac{1}{2}\|x\|_2^2$
- Fix some $\gamma \in (0, 1/(p - 2))$ (if $p = 2$, then $\gamma$ can be arbitrary positive)
- let $S_k$ be the ball of radius $r_k := k^\gamma$ and center 0
- We can take $L_k := (p - 1)2^{p-2}\|A\|^2 k^{\gamma(p-2)}$
- The convergence theorem ensures that $x_k$ converges to an optimal solution
- non-asymptotic rate of convergence is $O\left(\frac{1}{k^{1-\gamma(p-2)}}\right)$, namely, arbitrarily close to $O(1/k)$
A by-product: a general stability principle

Some assumptions:

\((X, d)\) is a metric space.

\(\emptyset \neq A \subseteq X\) is given.

\(h: X \to \mathbb{R}\) is uniformly continuous.

Notation:

The Hausdorff distance between \(A\) and \(\emptyset \neq A' \subseteq X\):

\[D_H(A, A') := \max \{\sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A)\}\]

where 

\[d(x, A) := \inf \{d(x, a) : a \in A\}\]

Roughly speaking, \(D_H(A, A')\) quantifies the metric similarity between \(A\) and \(A'\).
A by-product: a general stability principle

Some assumptions:

- $(X, d)$ is a metric space.
A by-product: a general stability principle

Some assumptions:

- $(X, d)$ is a metric space.
- $\emptyset \neq A \subseteq X$ is given.
A by-product: a general stability principle

Some assumptions:

- $(X, d)$ is a metric space.
- $\emptyset \neq A \subseteq X$ is given.
- $h : X \rightarrow \mathbb{R}$ is uniformly continuous.

Notation:
The Hausdorff distance between $A$ and $A' \subseteq X$:

$$D_H(A, A') := \max \{ \sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A) \}$$

where $d(x, A) := \inf \{ d(x, a) \mid a \in A \}$.

Roughly speaking, $D_H(A, A')$ quantifies the metric similarity between $A$ and $A'$. 

BISTA 25 March 2018 25 / 29
Some assumptions:

- \((X, d)\) is a metric space.
- \(\emptyset \neq A \subseteq X\) is given.
- \(h : X \to \mathbb{R}\) is uniformly continuous.

Notation:

- The Hausdorff distance between \(A\) and \(A'\): 
  \[D_H(A, A') := \max\{\sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A)\}\]

Roughly speaking, \(D_H(A, A')\) quantifies the metric similarity between \(A\) and \(A'\).
A by-product: a general stability principle

Some assumptions:

- $(X, d)$ is a metric space.
- $\emptyset \neq A \subseteq X$ is given.
- $h : X \to \mathbb{R}$ is uniformly continuous.

**Notation:** The Hausdorff distance between $A$ and $\emptyset \neq A' \subseteq X$: 

\[
D_H(A, A') := \max \left\{ \sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A) \right\},
\]

where $d(x, A) := \inf \{d(x, a) : a \in A\}$. Roughly speaking, $D_H(A, A')$ quantifies the metric similarity between $A$ and $A'$. 
A by-product: a general stability principle

Some assumptions:

- \((X, d)\) is a metric space.
- \(\emptyset \neq A \subseteq X\) is given.
- \(h : X \to \mathbb{R}\) is uniformly continuous.

Notation: The Hausdorff distance between \(A\) and \(\emptyset \neq A' \subseteq X\):

\[
D_H(A, A') := \max\{\sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A)\},
\]

Roughly speaking, \(D_H(A, A')\) quantifies the metric similarity between \(A\) and \(A'\).
A by-product: a general stability principle

Some assumptions:

- $(X, d)$ is a metric space.
- $\emptyset \neq A \subseteq X$ is given.
- $h : X \rightarrow \mathbb{R}$ is uniformly continuous.

Notation: The Hausdorff distance between $A$ and $\emptyset \neq A' \subseteq X$:

$$D_H(A, A') := \max\{\sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A)\},$$

where

$$d(x, A) := \inf\{d(x, a), a \in A\}.$$
Some assumptions:

- \((X, d)\) is a metric space.
- \(\emptyset \neq A \subseteq X\) is given.
- \(h : X \to \mathbb{R}\) is uniformly continuous.

Notation: The Hausdorff distance between \(A\) and \(\emptyset \neq A' \subseteq X\):

\[
D_H(A, A') := \max\{\sup_{a \in A} d(a, A'), \sup_{a' \in A'} d(a', A)\},
\]

where

\[
d(x, A) := \inf\{d(x, a), \ a \in A\}.
\]

Roughly speaking, \(D_H(A, A')\) quantifies the metric similarity between \(A\) and \(A'\).
The stability principle (intuitive formulation)

Principle

Consider an arbitrary uniformly continuous real function which is defined on a metric space. Then its extreme (optimal) values depend continuously on the subset over which they are computed. In other words, if we slightly change the subset, then the extreme values change slightly.

Remark

Uniform continuity is essential: there are counterexamples to the stability principle when the considered function is not uniformly continuous.
The stability principle (intuitive formulation)

Principle
Consider an arbitrary uniformly continuous real function which is defined on a metric space.

Remark
Uniform continuity is essential: there are counterexamples to the stability principle when the considered function is not uniformly continuous.
The stability principle (intuitive formulation)

Principle

Consider an arbitrary uniformly continuous real function which is defined on a metric space. Then its extreme (optimal) values depend continuously on the subset over which they are computed.
The stability principle (intuitive formulation)

Principle

Consider an arbitrary uniformly continuous real function which is defined on a metric space. Then its extreme (optimal) values depend continuously on the subset over which they are computed. In other words, if we slightly change the subset, then the extreme values change slightly.
The stability principle (intuitive formulation)

**Principle**

Consider an arbitrary uniformly continuous real function which is defined on a metric space. Then its extreme (optimal) values depend continuously on the subset over which they are computed. In other words, if we slightly change the subset, then the extreme values change slightly.

**Remark**

*Uniform continuity is essential*: there are counterexamples to the stability principle when the considered function is not uniformly continuous.
The stability principle (exact formulation)

Lemma

Under the previous mentioned assumptions, if $\sup_{a \in A} h(a) < \infty$, then for all $\epsilon > 0$, there exists $\delta > 0$ such that for each nonempty subset $A' \subseteq X$ satisfying $D_{\mathcal{H}}(A, A') < \delta$, we have $\sup_{a' \in A'} h(a') - \epsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \epsilon$.

If $\sup_{a \in A} h(a) = \infty$, then for all $M > 0$ there exists $\delta > 0$ such that for each nonempty subset $A' \subseteq X$ satisfying $D_{\mathcal{H}}(A, A') < \delta$, $M < \sup_{a' \in A'} h(a')$.

There are similar statements regarding $\inf_{a \in A} h(a)$.
Lemma

Under the previous mentioned assumptions, if \( \sup_{a \in A} h(a) < \infty \), then for all \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
\sup_{a' \in A'} h(a') - \epsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \epsilon,
\]

If \( \sup_{a \in A} h(a) = \infty \), then for all \( M > 0 \) there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \),

\[
M < \sup_{a' \in A'} h(a').
\]

There are similar statements regarding \( \inf_{a \in A} h(a) \).
The stability principle (exact formulation)

Lemma

Under the previous mentioned assumptions, if \( \sup_{a \in A} h(a) < \infty \), then for all \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
\sup_{a' \in A'} h(a') - \epsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \epsilon,
\]

If \( \sup_{a \in A} h(a) = \infty \), then for all \( M > 0 \) there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
M < \sup_{a' \in A'} h(a')
\]

There are similar statements regarding \( \inf_{a \in A} h(a) \).
The stability principle (exact formulation)

Lemma

Under the previous mentioned assumptions, if \( \sup_{a \in A} h(a) < \infty \), then for all \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
\sup_{a' \in A'} h(a') - \epsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \epsilon,
\]

If \( \sup_{a \in A} h(a) = \infty \), then for all \( M > 0 \) there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \),
Lemma

Under the previous mentioned assumptions, if \( \sup_{a \in A} h(a) < \infty \), then for all \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
\sup_{a' \in A'} h(a') - \varepsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \varepsilon,
\]

If \( \sup_{a \in A} h(a) = \infty \), then for all \( M > 0 \) there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \),

\[
M < \sup_{a' \in A'} h(a').
\]
The stability principle (exact formulation)

Lemma

Under the previous mentioned assumptions, if \( \sup_{a \in A} h(a) < \infty \), then for all \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
\sup_{a' \in A'} h(a') - \epsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \epsilon,
\]

If \( \sup_{a \in A} h(a) = \infty \), then for all \( M > 0 \) there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \),

\[
M < \sup_{a' \in A'} h(a').
\]

There are similar statements regarding \( \inf_{a \in A} h(a) \).
The stability principle (exact formulation)

Lemma

- Under the previous mentioned assumptions, if \( \sup_{a \in A} h(a) < \infty \), then for all \( \epsilon > 0 \), there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \), we have

\[
\sup_{a' \in A'} h(a') - \epsilon \leq \sup_{a \in A} h(a) \leq \sup_{a' \in A'} h(a') + \epsilon,
\]

- If \( \sup_{a \in A} h(a) = \infty \), then for all \( M > 0 \) there exists \( \delta > 0 \) such that for each nonempty subset \( A' \subseteq X \) satisfying \( D_H(A, A') < \delta \),

\[
M < \sup_{a' \in A'} h(a').
\]

- There are similar statements regarding \( \inf_{a \in A} h(a) \).

- In the case \( \sup_{a \in A} h(a) = \infty \) or \( \inf_{a \in A} h(a) = -\infty \), continuity suffices
A general scheme for nonconvex & nonsmooth optimization

Given:

- $h : X \to \mathbb{R}$ uniformly continuous,
- $\emptyset \neq C \subseteq X$

Goal: to estimate $\inf \{ h(x) : x \in C \}$

Suppose that we can approximate $C$ by a sequence $(C_k)_{k=1}^{\infty}$ of subsets of $X$ such that $\lim_{k \to \infty} D_H(C, C_k) = 0$

Assume that we are also able to compute an approximation $\tilde{s}_k$ to $\inf_{x \in C_k} h(x)$ so that $\lim_{k \to \infty} |\tilde{s}_k - \inf_{x \in C_k} h(x)| = 0$ (assuming that $\inf_{x \in C_k} h(x) \in \mathbb{R}$ for all $k \in \mathbb{N}$).

The method: compute $\tilde{s}_1$, $\tilde{s}_2$, etc.

The stability principle ensures that $\lim_{k \to \infty} \tilde{s}_k = \inf_{x \in C} h(x)$.

A common scenario:

- $X = \mathbb{R}^n$,
- $C \subset X$ is closed and bounded,
- $h$ is continuous on $C$ (or in a neighborhood of $C$),
- $C_k$ has a simple form (say, union of cubes).
A general scheme for nonconvex & nonsmooth optimization

- **Given:** \( h : X \rightarrow \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)
A general scheme for nonconvex & nonsmooth optimization

- **Given**: \( h : X \rightarrow \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)
- **Goal**: to estimate\( \inf_{x \in C} h(x) \)
A general scheme for nonconvex & nonsmooth optimization

- **Given**: \( h : X \rightarrow \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)
- **Goal**: to estimate
  \[
  \inf \{ h(x) : x \in C \}
  \]
A general scheme for nonconvex & nonsmooth optimization

- **Given**: \( h : X \to \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)
- **Goal**: to estimate
  \[ \inf \{ h(x) : x \in C \} \]
- Suppose that we can approximate \( C \) by a sequence \( (C_k)_{k=1}^\infty \) of subsets of \( X \) such that \( \lim_{k \to \infty} D_H(C, C_k) = 0 \)

A common scenario:
\( X = \mathbb{R}^n, C \subseteq X \) is closed and bounded, \( h \) is continuous on \( C \) (or in a neighborhood of \( C \)), \( C_k \) has a simple form (say, union of cubes).
A general scheme for nonconvex & nonsmooth optimization

- **Given**: \( h : X \rightarrow \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)
- **Goal**: to estimate
  \[
  \inf \{ h(x) : x \in C \}
  \]

Suppose that we can approximate \( C \) by a sequence \( (C_k)_{k=1}^{\infty} \) of subsets of \( X \) such that \( \lim_{k \to \infty} D_H(C, C_k) = 0 \)

Assume that we are also able to compute an approximation \( \tilde{s}_k \) to \( \inf_{x \in C_k} h(x) \) so that \( \lim_{k \to \infty} |\tilde{s}_k - \inf_{x \in C_k} h(x)| = 0 \) (assuming that \( \inf_{x \in C_k} h(x) \in \mathbb{R} \) for all \( k \in \mathbb{N} \)).
A general scheme for nonconvex & nonsmooth optimization

- **Given:** \( h : X \rightarrow \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)

- **Goal:** to estimate \( \inf \{ h(x) : x \in C \} \)

Suppose that we can approximate \( C \) by a sequence \( (C_k)_{k=1}^{\infty} \) of subsets of \( X \) such that \( \lim_{k \to \infty} D_H(C, C_k) = 0 \)

Assume that we are also able to compute an approximation \( \tilde{s}_k \) to \( \inf_{x \in C_k} h(x) \) so that \( \lim_{k \to \infty} |\tilde{s}_k - \inf_{x \in C_k} h(x)| = 0 \) (assuming that \( \inf_{x \in C_k} h(x) \in \mathbb{R} \) for all \( k \in \mathbb{N} \)).

**The method:** compute \( \tilde{s}_1, \tilde{s}_2, \) etc.
A general scheme for nonconvex & nonsmooth optimization

- **Given:** \( h : X \rightarrow \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)
- **Goal:** to estimate 
  \( \inf \{ h(x) : x \in C \} \)

Suppose that we can approximate \( C \) by a sequence \( (C_k)_{k=1}^{\infty} \) of subsets of \( X \) such that \( \lim_{k \rightarrow \infty} D_H(C, C_k) = 0 \)

Assume that we are also able to compute an approximation \( \tilde{s}_k \) to \( \inf_{x \in C_k} h(x) \) so that \( \lim_{k \rightarrow \infty} |\tilde{s}_k - \inf_{x \in C_k} h(x)| = 0 \) (assuming that \( \inf_{x \in C_k} h(x) \in \mathbb{R} \) for all \( k \in \mathbb{N} \)).

- **The method:** compute \( \tilde{s}_1, \tilde{s}_2, \) etc.
- **the stability principle ensures** that \( \lim_{k \rightarrow \infty} \tilde{s}_k = \inf_{x \in C} h(x) \).
A general scheme for nonconvex & nonsmooth optimization

- **Given**: \( h : X \to \mathbb{R} \) uniformly continuous, \( \emptyset \neq C \subseteq X \)

- **Goal**: to estimate \( \inf \{ h(x) : x \in C \} \)

Suppose that we can approximate \( C \) by a sequence \((C_k)_{k=1}^{\infty}\) of subsets of \( X \) such that \( \lim_{k \to \infty} D_H(C, C_k) = 0 \)

Assume that we are also able to compute an approximation \( \tilde{s}_k \) to \( \inf_{x \in C_k} h(x) \) so that \( \lim_{k \to \infty} |\tilde{s}_k - \inf_{x \in C_k} h(x)| = 0 \) (assuming that \( \inf_{x \in C_k} h(x) \in \mathbb{R} \) for all \( k \in \mathbb{N} \)).

- **The method**: compute \( \tilde{s}_1, \tilde{s}_2, \) etc.

- **the stability principle ensures** that \( \lim_{k \to \infty} \tilde{s}_k = \inf_{x \in C} h(x) \).

- **a common scenario**: \( X = \mathbb{R}^n \), \( C \subset X \) is closed and bounded, \( h \) is continuous on \( C \) (or in a neighborhood of \( C \)), \( C_k \) has a simple form (say, union of cubes).
The End

P.S. The slides can be found online:

http://w3.impa.br/~dream/talks