Relative Smoothness: New Paradigm in Convex Optimization

Yurii Nesterov (CORE/INMA, UCL, Belgium)

September 4, 2019 (EUSIPCO 2019, A Coruña, Spain)

Traditional frameworks

Traditional frameworks

NB:

Traditional frameworks

NB: They are based on implementable *operations*.

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function*

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

2. *Euclidean projection*

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

2. *Euclidean projection* (mid-60's)

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

2. *Euclidean projection* (mid-60's)

3. *Minimization of the augmented linear model of objective function*

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

2. *Euclidean projection* (mid-60's)

3. *Minimization of the augmented linear model of objective function*

(Mirror-descent methods,

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

2. *Euclidean projection* (mid-60's)

3. *Minimization of the augmented linear model of objective function*

(Mirror-descent methods, Nemirovsky-Yudin, mid-70's)

Traditional frameworks

NB: They are based on implementable *operations*.

Examples:

1. *Minimization of linear function* (Frank-Wolfe, 1956)

2. *Euclidean projection* (mid-60's)

3. *Minimization of the augmented linear model of objective function*

(Mirror-descent methods, Nemirovsky-Yudin, mid-70's)

Not too many possibilities for development of minimization methods.

New possibilities

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

Main idea:

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

Main idea:

1. Assume that we know how to solve some *simple* minimization problems.

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

Main idea:

1. Assume that we know how to solve some *simple* minimization problems.

2. Then we can use this *know-how* for solving more complicated problems.

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

Main idea:

1. Assume that we know how to solve some *simple* minimization problems.

2. Then we can use this *know-how* for solving more complicated problems.

3. For that, the objective function of our complicated problems must be *similar* to the initial simple functions.

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

Main idea:

1. Assume that we know how to solve some *simple* minimization problems.

2. Then we can use this *know-how* for solving more complicated problems.

3. For that, the objective function of our complicated problems must be *similar* to the initial simple functions.

Main question:

New possibilities

They appear in the framework of *Relative Smoothness Condition*.

Main idea:

1. Assume that we know how to solve some *simple* minimization problems.

2. Then we can use this *know-how* for solving more complicated problems.

3. For that, the objective function of our complicated problems must be *similar* to the initial simple functions.

Main question: How to measure this *similarity*?

Relative Smoothness Condition

Relative Smoothness Condition

Let $d \in \operatorname{dom} d \subseteq \mathbb{R}^n$ be a *simple convex function*.

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC)

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$,

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \quad \preceq \quad L \nabla^2 d(x),$$

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \quad \preceq \quad L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x)$$

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \le \mu \le L$, such that

$$\nabla^2 f(x) \preceq L \nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu \nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L \nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu \nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant:

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L \nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu \nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d,

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

Relative Smoothness Condition

Let $d \in \operatorname{dom} d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\operatorname{dom} d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \operatorname{dom} d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle$$

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \quad \preceq \quad L \nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \quad \succeq \quad \mu \nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x,y) \quad = \quad d(y) - d(x) - \langle \nabla d(x), y - x \rangle \; \geq \; 0, \quad x, y \in \mathrm{dom}\, d.$$

Relative Smoothness Condition

Let $d \in \operatorname{dom} d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\operatorname{dom} d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \operatorname{dom} d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x\rangle \geq 0, \quad x, y \in \operatorname{dom} d.$$

Then f satisfies RSC if

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \le \mu \le L$, such that

$$\nabla^2 f(x) \preceq L \nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu \nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \ge 0, \quad x, y \in \mathrm{dom}\, d.$$

Then f satisfies RSC if

$$\mu \beta_d(x,y) \le \beta_f(x,y) \le L \beta_d(x,y), \quad x, y \in \mathrm{dom}\, d.$$

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq 0, \quad x, y \in \mathrm{dom}\, d.$$

Then f satisfies RSC if

$$\mu\beta_d(x, y) \leq \beta_f(x, y) \leq L\beta_d(x, y), \quad x, y \in \mathrm{dom}\, d.$$

Main advantage:

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L \nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu \nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq 0, \quad x, y \in \mathrm{dom}\, d.$$

Then f satisfies RSC if

$$\mu \beta_d(x, y) \leq \beta_f(x, y) \leq L \beta_d(x, y), \quad x, y \in \mathrm{dom}\, d.$$

Main advantage:

If we can easily minimize d,

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \le \mu \le L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x\rangle \ge 0, \quad x, y \in \mathrm{dom}\, d.$$

Then f satisfies RSC if

$$\mu\beta_d(x,y) \le \beta_f(x,y) \le L\beta_d(x,y), \quad x, y \in \mathrm{dom}\, d.$$

Main advantage:

If we can easily minimize d, then we can minimize f very efficiently

Relative Smoothness Condition

Let $d \in \mathrm{dom}\, d \subseteq \mathbb{R}^n$ be a *simple convex function*.

We say that function $f \in \mathbb{C}^2(\mathrm{dom}\, d)$ satisfies *Relative Smoothness Condition* (RSC) if there exist constants μ and L, $0 \leq \mu \leq L$, such that

$$\nabla^2 f(x) \preceq L\nabla^2 d(x), \quad \text{(Bauschke, Beck, Teboulle, 2017)},$$

$$\nabla^2 f(x) \succeq \mu\nabla^2 d(x) \quad \text{(Lu, Freund, N., 2018)}$$

for all $x \in \mathrm{dom}\, d$.

First-order variant: For function d, define the *Bregmann distance*:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq 0, \quad x, y \in \mathrm{dom}\, d.$$

Then f satisfies RSC if

$$\mu\beta_d(x,y) \leq \beta_f(x,y) \leq L\beta_d(x,y), \quad x, y \in \mathrm{dom}\, d.$$

Main advantage:

If we can easily minimize d, then we can minimize f very efficiently

by the simple *Gradient Schemes*.

Simple Gradient Method

Simple Gradient Method

Problem:

Simple Gradient Method

Problem: $\min_{x \in Q} f(x),$

Simple Gradient Method

Problem: $\min_{x \in Q} f(x),$

where

Simple Gradient Method

Problem: $$\min_{x \in Q} f(x),$$

where

- Q is a closed convex set.

Simple Gradient Method

Problem: $\min_{x \in Q} f(x),$

where

- Q is a closed convex set.
- f satisfies RSC on Q

Simple Gradient Method

Problem: $\min_{x \in Q} f(x),$

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \operatorname{dom} d$.

Simple Gradient Method

Problem: $\min_{x \in Q} f(x),$

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \operatorname{dom} d$.

Main Assumption:

Simple Gradient Method

Problem: $\min_{x \in Q} f(x)$,

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \mathrm{dom}\, d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{\langle g, x \rangle + d(x)\}}$

Simple Gradient Method

Problem: $\min_{x \in Q} f(x)$,

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \operatorname{dom} d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{\langle g, x\rangle + d(x)\}}$

with any vector $g \in \mathbb{R}^n$.

Simple Gradient Method

Problem: $\min_{x \in Q} f(x),$

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \operatorname{dom} d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{\langle g, x \rangle + d(x)\}}$

with any vector $g \in \mathbb{R}^n$.

Gradient Method:

Simple Gradient Method

Problem: $\min_{x \in Q} f(x)$,

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \mathrm{dom}\, d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{ \langle g, x \rangle + d(x) \}}$

with any vector $g \in \mathbb{R}^n$.

Gradient Method: Choose $x_0 \in Q$ and iterate

Simple Gradient Method

Problem: $\min_{x \in Q} f(x)$,

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \mathrm{dom}\, d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{\langle g, x \rangle + d(x)\}}$

with any vector $g \in \mathbb{R}^n$.

Gradient Method: Choose $x_0 \in Q$ and iterate

$$x_{k+1} \quad \in \quad \operatorname{Arg}\min_{x \in Q} \Big\{ \langle \nabla f(x_k), x - x_k \rangle + L\beta_d(x_k, x) \Big\}, \quad k \geq 0.$$

Simple Gradient Method

Problem: $\min_{x \in Q} f(x)$,

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \operatorname{dom} d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{ \langle g, x \rangle + d(x) \}}$

with any vector $g \in \mathbb{R}^n$.

Gradient Method: Choose $x_0 \in Q$ and iterate

$$x_{k+1} \quad \in \quad \operatorname{Arg}\min_{x \in Q} \Big\{ \langle \nabla f(x_k), x - x_k \rangle + L \beta_d(x_k, x) \Big\}, \quad k \geq 0.$$

Simple property:

Simple Gradient Method

Problem: $\min_{x \in Q} f(x)$,

where

- Q is a closed convex set.
- f satisfies RSC on Q for some scaling function d with $Q \subseteq \mathrm{dom}\, d$.

Main Assumption: we can solve the problem $\boxed{\min_{x \in Q} \{\langle g, x \rangle + d(x)\}}$

with any vector $g \in \mathbb{R}^n$.

Gradient Method: Choose $x_0 \in Q$ and iterate

$$x_{k+1} \in \operatorname{Arg}\min_{x \in Q} \Big\{ \langle \nabla f(x_k), x - x_k \rangle + L\beta_d(x_k, x) \Big\}, \quad k \geq 0.$$

Simple property: $f(x_{k+1}) \leq f(x_k)$, $k \geq 0$.

Global rate of convergence

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number).

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process,

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB:

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\ell_T(x) \leq f(x)$ for all x in Q.

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\quad \ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

$$f(x_T) + L\beta_d(x_T, x) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x) + \ell_T(x).$$

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\quad \ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

$$f(x_T) + L\beta_d(x_T, x) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x) + \ell_T(x).$$

Corrollary:

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

$$f(x_T) + L\beta_d(x_T, x) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x) + \ell_T(x).$$

Corrollary: $f(x_T) - f^* \leq \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x^*)$.

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

$$f(x_T) + L\beta_d(x_T, x) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x) + \ell_T(x).$$

Corrollary: $f(x_T) - f^* \leq \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x^*)$.

NB:

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

$$f(x_T) + L\beta_d(x_T, x) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x) + \ell_T(x).$$

Corrollary: $f(x_T) - f^* \leq \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x^*)$.

NB: When $\mu \to 0$, we get the following limit:

Global rate of convergence

Denote $\gamma = \frac{\mu}{L}$ (Condition number). We assume that μ and L are known.

During the process, for $T \geq 1$, we update the following *model* of the objective function:

$$\ell_T(x) = \frac{\gamma}{1-(1-\gamma)^T} \sum_{k=1}^{T} (1-\gamma)^{T-k} \Big[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \mu \beta_d(x_k, x) \Big].$$

NB: $\ell_T(x) \leq f(x)$ for all x in Q.

Theorem. For all $T \geq 1$ and $x \in Q$ we have

$$f(x_T) + L\beta_d(x_T, x) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x) + \ell_T(x).$$

Corrollary: $f(x_T) - f^* \leq \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \beta_d(x_0, x^*)$.

NB: When $\mu \to 0$, we get the following limit:

$$f(x_T) - f^* \leq \frac{L}{T} \beta_d(x_0, x^*).$$

Accuracy certificate

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

$$f(x_T) - \ell_T^*(\mathcal{N}) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

$$f(x_T) - \ell_T^*(\mathcal{N}) \quad \leq \quad \tfrac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Most natural choice:

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x)$.

At the same time,

$$f(x_T) - \ell_T^*(\mathcal{N}) \quad \leq \quad \tfrac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Most natural choice:

$$\mathcal{N} = \{x : \ \beta_d(x_0, x) \leq D\}.$$

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

$$f(x_T) - \ell_T^*(\mathcal{N}) \quad \leq \quad \tfrac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Most natural choice:

$$\mathcal{N} = \{x : \ \beta_d(x_0, x) \leq D\}.$$

NB:

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\rm def}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

$$f(x_T) - \ell_T^*(\mathcal{N}) \quad \leq \quad \frac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Most natural choice:

$$\mathcal{N} = \{x : \ \beta_d(x_0, x) \leq D\}.$$

NB: 1. In many situations, $\mathcal{N}$ and D are explicitly known.

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell^*_T(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

$$f(x_T) - \ell^*_T(\mathcal{N}) \quad \leq \quad \tfrac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Most natural choice:

$$\mathcal{N} = \{x : \; \beta_d(x_0, x) \leq D\}.$$

NB: 1. In many situations, $\mathcal{N}$ and D are explicitly known.

2. The rate of convergence is not as good as for the Fast Gradient Methods.

Accuracy certificate

Assume that we know some set $\mathcal{N}$ containing the solution x^*.

Then $\quad f^* \geq \ell_T^*(\mathcal{N}) \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q \bigcap \mathcal{N}} \ell_T(x).$

At the same time,

$$f(x_T) - \ell_T^*(\mathcal{N}) \quad \leq \quad \tfrac{\mu(1-\gamma)^T}{1-(1-\gamma)^T} \max_{x \in Q \bigcap \mathcal{N}} \beta_d(x_0, x).$$

Most natural choice:

$$\mathcal{N} = \{x : \; \beta_d(x_0, x) \leq D\}.$$

NB: 1. In many situations, $\mathcal{N}$ and D are explicitly known.

2. The rate of convergence is not as good as for the Fast Gradient Methods.

However, the field of applications is much wider.

Examples

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$,

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*.

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|_{(2)}^r)$.

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$,

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|_{(2)}^r)$.

Then for $d(x) = \frac{1}{r+2}\|x\|_{(2)}^{r+2} + \frac{1}{2}\|x\|_{(2)}^2$, function f satisfies RSC with L.

3. *Splitting the objective*.

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x)$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|_{(2)}^r)$.

Then for $d(x) = \frac{1}{r+2}\|x\|_{(2)}^{r+2} + \frac{1}{2}\|x\|_{(2)}^2$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|_{(2)}^r)$.

Then for $d(x) = \frac{1}{r+2}\|x\|_{(2)}^{r+2} + \frac{1}{2}\|x\|_{(2)}^2$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \le L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians*

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$ with Laplacian $A \succeq 0$,

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$ with Laplacian $A \succeq 0$, we represent

$$\langle Ax, x \rangle = \sum_{i=1}^{m} \langle a_i, x \rangle^2$$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$ with Laplacian $A \succeq 0$, we represent

$$\langle Ax, x \rangle = \sum_{i=1}^{m} \langle a_i, x \rangle^2 \;=\; \sum_{i \in T} \langle a_i, x \rangle^2 + \sum_{i \notin T} \langle a_i, x \rangle^2$$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|_{(2)}^2$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|_{(2)}^r)$.

Then for $d(x) = \frac{1}{r+2}\|x\|_{(2)}^{r+2} + \frac{1}{2}\|x\|_{(2)}^2$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$ with Laplacian $A \succeq 0$, we represent

$$\langle Ax, x\rangle = \sum_{i=1}^{m} \langle a_i, x\rangle^2 \; = \; \sum_{i\in T} \langle a_i, x\rangle^2 + \sum_{i\notin T} \langle a_i, x\rangle^2 \; \stackrel{\mathrm{def}}{=} \; B + C,$$

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$ with Laplacian $A \succeq 0$, we represent

$$\langle Ax, x\rangle = \sum_{i=1}^{m} \langle a_i, x\rangle^2 \;=\; \sum_{i\in T} \langle a_i, x\rangle^2 + \sum_{i\notin T} \langle a_i, x\rangle^2 \stackrel{\text{def}}{=} B + C,$$

with $C \preceq LB$,

Examples

1. *Strongly convex functions* with Lipschitz-continuous gradient.

Indeed, choosing $d(x) = \frac{1}{2}\|x\|^2_{(2)}$, we have

$$\mu I \preceq \nabla^2 f(x) \preceq LI \text{ for all } x \in \mathbb{R}^n.$$

2. *Bounded growth of the Hessian*. Let $\|\nabla^2 f(x)\| \leq L(1 + \|x\|^r_{(2)})$.

Then for $d(x) = \frac{1}{r+2}\|x\|^{r+2}_{(2)} + \frac{1}{2}\|x\|^2_{(2)}$, function f satisfies RSC with L.

3. *Splitting the objective*. Let $F(x) = f(x) + \psi(x)$ with *simple* $\psi(\cdot)$ and

$$\nabla^2 f(x) \preceq \gamma \nabla^2 \psi(x).$$

Then $\mu = 1$ and $L = 1 + \gamma$.

Compare: *Fast inversion of Laplacians* (Spilmann, Tao (2010), ...)

For solving the system $Ax = b$ with Laplacian $A \succeq 0$, we represent

$$\langle Ax, x\rangle = \sum_{i=1}^m \langle a_i, x\rangle^2 \;=\; \sum_{i\in T}\langle a_i, x\rangle^2 + \sum_{i\notin T}\langle a_i, x\rangle^2 \;\stackrel{\text{def}}{=}\; B + C,$$

with $C \preceq LB$, and use B as a *preconditioner*.

Example: Third-order method

Example: Third-order method

Problem:

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Example: Third-order method

Problem: $\min\limits_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

Example: Third-order method

Problem: $\min\limits_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned}\Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2}\langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &+ \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24}\|x - \bar{x}\|_{(2)}^4.\end{aligned}$$

Example: Third-order method

Problem: $\min\limits_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\Omega_{\bar{x},H}(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle$$
$$+ \frac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \frac{H}{24} \| x - \bar{x} \|_{(2)}^4.$$

Theorem (N.2018).

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\Omega_{\bar{x},H}(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{1}{2}\langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle$$
$$+ \frac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \frac{H}{24}\|x - \bar{x}\|_{(2)}^4.$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Example: Third-order method

Problem: $\min\limits_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}^{3,1}_L(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned} \Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ & \quad + \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|^4_{(2)}. \end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question:

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned} \Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &+ \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4. \end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Example: Third-order method

Problem: $\min\limits_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned}\Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &+ \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4.\end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer:

Example: Third-order method

Problem: $\min\limits_{x \in \mathbb{R}^n} f(x),$

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned}\Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &\quad + \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4.\end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned}\Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &\quad + \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4.\end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Let $\tau = \sqrt{\frac{H}{3L}} > 1$.

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\Omega_{\bar{x},H}(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle$$

$$+ \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4.$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Let $\tau = \sqrt{\frac{H}{3L}} > 1$. Define

$$d_{\bar{x},\tau}(x) = \tfrac{1}{2} \left(1 - \tfrac{1}{\tau}\right) \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle + \tfrac{\tau(\tau-1)L}{8} \| x - \bar{x} \|_{(2)}^4.$$

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned}\Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &\quad + \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4.\end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Let $\tau = \sqrt{\frac{H}{3L}} > 1$. Define

$$d_{\bar{x},\tau}(x) = \tfrac{1}{2}\left(1 - \tfrac{1}{\tau}\right) \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle + \tfrac{\tau(\tau-1)L}{8} \| x - \bar{x} \|_{(2)}^4.$$

Then $\nabla^2 d_{\bar{x},\tau}(x) \preceq \nabla \Omega_{\bar{x},H}(x) \preceq \frac{\tau+1}{\tau-1} \nabla^2 d_{\bar{x},\tau}(x)$.

Example: Third-order method

Problem: $\min_{x\in\mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned}\Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x-\bar{x}\rangle + \tfrac{1}{2}\langle \nabla^2 f(\bar{x})(x-\bar{x}), x-\bar{x}\rangle \\ &\quad + \tfrac{1}{6} D^3 f(\bar{x})[y-\bar{x}]^3 + \tfrac{H}{24}\|x-\bar{x}\|_{(2)}^4.\end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Let $\tau = \sqrt{\frac{H}{3L}} > 1$. Define

$$d_{\bar{x},\tau}(x) = \tfrac{1}{2}\left(1-\tfrac{1}{\tau}\right)\langle \nabla^2 f(\bar{x})(x-\bar{x}), x-\bar{x}\rangle + \tfrac{\tau(\tau-1)L}{8}\|x-\bar{x}\|_{(2)}^4.$$

Then $\nabla^2 d_{\bar{x},\tau}(x) \preceq \nabla\Omega_{\bar{x},H}(x) \preceq \frac{\tau+1}{\tau-1}\nabla^2 d_{\bar{x},\tau}(x)$.

Choosing $\tau = 2$

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\begin{aligned} \Omega_{\bar{x},H}(x) &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle \\ &\quad + \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|_{(2)}^4 . \end{aligned}$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Let $\tau = \sqrt{\frac{H}{3L}} > 1$. Define

$$d_{\bar{x},\tau}(x) = \tfrac{1}{2}\left(1 - \tfrac{1}{\tau}\right) \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle + \tfrac{\tau(\tau-1)L}{8} \| x - \bar{x} \|_{(2)}^4 .$$

Then $\nabla^2 d_{\bar{x},\tau}(x) \preceq \nabla \Omega_{\bar{x},H}(x) \preceq \frac{\tau+1}{\tau-1} \nabla^2 d_{\bar{x},\tau}(x)$.

Choosing $\tau = 2$ (this is $H = 12L$),

Example: Third-order method

Problem: $\min_{x \in \mathbb{R}^n} f(x)$,

where function $f \in \mathbb{C}_L^{3,1}(\mathbb{R}^n)$ is convex.

Consider the regularized Taylor polynomial of degree three:

$$\Omega_{\bar{x},H}(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle$$

$$+ \tfrac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \tfrac{H}{24} \| x - \bar{x} \|^4_{(2)}.$$

Theorem (N.2018). If $H \geq 3L$ then $\Omega_{\bar{x},H}(\cdot)$ is a convex polynomial.

Main question: How to minimize it?

Answer: by the Gradient Method based on RSC.

Let $\tau = \sqrt{\frac{H}{3L}} > 1$. Define

$$d_{\bar{x},\tau}(x) = \tfrac{1}{2} \left(1 - \tfrac{1}{\tau}\right) \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle + \tfrac{\tau(\tau - 1)L}{8} \| x - \bar{x} \|^4_{(2)}.$$

Then $\nabla^2 d_{\bar{x},\tau}(x) \preceq \nabla \Omega_{\bar{x},H}(x) \preceq \frac{\tau+1}{\tau-1} \nabla^2 d_{\bar{x},\tau}(x)$.

Choosing $\tau = 2$ (this is $H = 12L$), we have $\gamma = \frac{1}{3}$.

Nondifferentiable Functions

Nondifferentiable Functions

Main diffculties:

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing,

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

if for any $x, y \in \mathrm{dom}\, f$

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

if for any $x, y \in \mathrm{dom}\, f$ and $g_f \in \partial f(x)$,

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

if for any $x, y \in \mathrm{dom}\, f$ and $g_f \in \partial f(x)$, we have

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\, \varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

if for any $x, y \in \mathrm{dom}\, f$ and $g_f \in \partial f(x)$, we have

$$\langle g_f, x - y \rangle \leq \varphi^{-1}(\beta_d(x, y)),$$

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

if for any $x, y \in \mathrm{dom}\, f$ and $g_f \in \partial f(x)$, we have

$$\langle g_f, x - y \rangle \leq \varphi^{-1}(\beta_d(x, y)),$$

where $\varphi^{-1}(\cdot)$ is the inverse of function $\varphi(\cdot)$,

Nondifferentiable Functions

Main diffculties:

- The subgradients are discontinuous.
- The norm of subgradient does not describe well the growth of the function.

Definition.

Let function φ be strictly increasing, $0 \in \mathrm{dom}\,\varphi \subseteq \mathbb{R}$, and $\varphi(0) = 0$.

We say that f has a *Bounded Growth* with respect to $d(\cdot)$ and $\varphi(\cdot)$

if for any $x, y \in \mathrm{dom}\, f$ and $g_f \in \partial f(x)$, we have

$$\langle g_f, x - y \rangle \leq \varphi^{-1}(\beta_d(x, y)),$$

where $\varphi^{-1}(\cdot)$ is the inverse of function $\varphi(\cdot)$, and

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \quad x, y \in \mathrm{dom}\, d.$$

Examples

Examples

1. *Functions with bounded subgradients*.

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle$$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x-y\|^2,\ x,y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \operatorname{dom} d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \operatorname{dom} f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\|$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x,y)}$.

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x-y\|^2, \; x, y \in \operatorname{dom} d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \operatorname{dom} f,$$

we have $\langle g, x - y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition.

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x-y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x, y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y-x \rangle \geq \tfrac{1}{2}\|x-y\|^2, \; x,y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x-y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms:

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x, y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h) \rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h \rangle^{1/2}$,

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y-x \rangle \geq \tfrac{1}{2}\|x-y\|^2,\ x,y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x-y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x\rangle \geq \frac{1}{2}\|x-y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y\rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Def.

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x-y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous*

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x-y\|^2, \; x,y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x-y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h) \rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d)

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x-y\|^2, \; x, y \in \operatorname{dom} d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \operatorname{dom} f,$$

we have $\langle g, x - y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h) \rangle^{1/2} \text{ for all } x \in \operatorname{dom} d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d) if

$$\|g\|_x^* \leq M \text{ for all } g \in \partial f(x),$$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x - y\|^2,\ x, y \in \operatorname{dom} d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \operatorname{dom} f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x) h, h) \rangle^{1/2} \text{ for all } x \in \operatorname{dom} d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x) h, h \rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g \rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d) if

$$\|g\|_x^* \leq M \text{ for all } g \in \partial f(x),\ x \in \operatorname{dom} f.$$

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x-y\|^2,\ x,y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d) if

$$\|g\|_x^* \leq M \text{ for all } g \in \partial f(x),\ x \in \mathrm{dom}\, f.$$

Theorem:

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h) \rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h \rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g \rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d) if

$$\|g\|_x^* \leq M \text{ for all } g \in \partial f(x), \; x \in \mathrm{dom}\, f.$$

Theorem: function f has Bounded Growth

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \frac{1}{2}\|x - y\|^2, \; x, y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x - y \rangle \leq M\|x - y\| \leq M\sqrt{2\beta_d(x, y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x)h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x)h, h\rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g\rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d) if

$$\|g\|_x^* \leq M \text{ for all } g \in \partial f(x), \; x \in \mathrm{dom}\, f.$$

Theorem: function f has Bounded Growth with $\varphi(\tau) = \omega\left(\frac{\tau}{M}\right)$,

Examples

1. *Functions with bounded subgradients*. Assume that the scaling function is strongly convex:

$$\beta_d(x,y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \geq \tfrac{1}{2}\|x-y\|^2,\ x,y \in \mathrm{dom}\, d.$$

Then for any function with bounded subgradients:

$$\|g\|_* \leq M \text{ for all } g \in \partial f(x) \text{ and all } x \in \mathrm{dom}\, f,$$

we have $\langle g, x-y \rangle \leq M\|x-y\| \leq M\sqrt{2\beta_d(x,y)}$.

Thus, we can take $\varphi(\tau) = \frac{\tau^2}{2M^2}$.

2. Local Lipschitz condition. Let $d(\cdot)$ be a *self-concordant function*:

$$D^3 d(x)[h]^3 \leq 2\langle \nabla^2 d(x) h, h)\rangle^{1/2} \text{ for all } x \in \mathrm{dom}\, d \text{ and } h \in \mathbb{R}^n.$$

Local norms: $\|h\|_x = \langle \nabla^2 f(x) h, h \rangle^{1/2}$, $\|g\|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g \rangle^{1/2}$.

Def. We call f *Locally Lipschitz Continuous* (with respect to d) if

$$\|g\|_x^* \leq M \text{ for all } g \in \partial f(x),\ x \in \mathrm{dom}\, f.$$

Theorem: function f has Bounded Growth with $\varphi(\tau) = \omega\left(\frac{\tau}{M}\right)$, where $\omega(\tau) = \tau - \ln(1+\tau)$.

Examples

Examples

3. *Positive concave functions*.

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set,

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q with

$$M = 1.$$

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q with

$$M = 1.$$

NB:

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q with

$$M = 1.$$

NB: By minimizing $f(\cdot)$

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q with

$$M = 1.$$

NB: By minimizing $f(\cdot)$ with *absolute accuracy*,

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q with

$$M = 1.$$

NB: By minimizing $f(\cdot)$ with *absolute accuracy*, we maximize $\psi(\cdot)$

Examples

3. *Positive concave functions*.

Consider the following *maximization* problem:

$$\max_{x \in Q} \psi(x),$$

where

- Q is a closed bounded convex set, for which we know a *self-concordant barrier* $d(\cdot)$,
- $\psi(\cdot)$ is a *concave function*.

Assumption. Function $\psi(\cdot)$ is *non-negative* on Q.

Theorem. Function $f(x) = -\ln \psi(x)$ has Bounded Growth on Q with

$$M = 1.$$

NB: By minimizing $f(\cdot)$ with *absolute accuracy*, we maximize $\psi(\cdot)$ with *relative accuracy*.

Primal Gradient Method

Primal Gradient Method

Consider the problem

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \frac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method:

Primal Gradient Method

Consider the problem $\boxed{\min_{x\in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x\in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h}\beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k),\ k \geq 0,}$

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \; k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \ k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem.

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \; k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem. Denote $S_T = \sum\limits_{k=0}^{T} h_k$.

Primal Gradient Method

Consider the problem $\boxed{\min_{x\in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x\in Q} \left\{ \langle g, x - \bar{x}\rangle + \tfrac{1}{h}\beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \; k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem. Denote $S_T = \sum\limits_{k=0}^{T} h_k$. Then

$$\Delta_T \stackrel{\mathrm{def}}{=} \min_{0\leq k\leq T} f(x_k) - f^* \leq \tfrac{1}{S_T}\left[\beta_d(x_0, x^*) + \sum_{k=0}^{T} \varphi_*(h_k)\right],$$

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \; k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem. Denote $S_T = \sum\limits_{k=0}^{T} h_k$. Then

$$\Delta_T \stackrel{\mathrm{def}}{=} \min_{0 \leq k \leq T} f(x_k) - f^* \leq \tfrac{1}{S_T} \left[\beta_d(x_0, x^*) + \sum_{k=0}^{T} \varphi_*(h_k) \right],$$

where $\varphi_*(t) = \max\limits_{\tau \geq 0} \{ \tau t - \varphi(\tau) \}$.

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \frac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \; k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem. Denote $S_T = \sum_{k=0}^{T} h_k$. Then

$$\Delta_T \stackrel{\mathrm{def}}{=} \min_{0 \leq k \leq T} f(x_k) - f^* \leq \frac{1}{S_T} \left[\beta_d(x_0, x^*) + \sum_{k=0}^{T} \varphi_*(h_k) \right],$$

where $\varphi_*(t) = \max_{\tau \geq 0} \{ \tau t - \varphi(\tau) \}$.

Examples:

Primal Gradient Method

Consider the problem $\boxed{\min_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \ k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem. Denote $S_T = \sum\limits_{k=0}^{T} h_k$. Then

$$\Delta_T \overset{\text{def}}{=} \min_{0 \leq k \leq T} f(x_k) - f^* \leq \tfrac{1}{S_T} \left[\beta_d(x_0, x^*) + \sum_{k=0}^{T} \varphi_*(h_k) \right],$$

where $\varphi_*(t) = \max\limits_{\tau \geq 0} \{ \tau t - \varphi(\tau) \}$.

Examples: $h_k \approx \frac{1}{\sqrt{k+1}}$,

Primal Gradient Method

Consider the problem $\boxed{\min\limits_{x \in Q} f(x),}$ where

- Q is a closed convex set,
- function f has Bounded Growth with respect to (d, φ).

Define the *Bregmann mapping*

$$T_h(\bar{x}, g) = \arg\min_{x \in Q} \left\{ \langle g, x - \bar{x} \rangle + \tfrac{1}{h} \beta_d(\bar{x}, x) \right\}.$$

Gradient Method: $\boxed{x_{k+1} = T_{h_k}(x_k, g_k), \ k \geq 0,}$

where $g_k \in \partial f(x_k)$, $h_k > 0$, and $x_0 \in Q$.

Theorem. Denote $S_T = \sum\limits_{k=0}^{T} h_k$. Then

$$\Delta_T \stackrel{\mathrm{def}}{=} \min_{0 \leq k \leq T} f(x_k) - f^* \leq \tfrac{1}{S_T} \left[\beta_d(x_0, x^*) + \sum_{k=0}^{T} \varphi_*(h_k) \right],$$

where $\varphi_*(t) = \max\limits_{\tau \geq 0} \{ \tau t - \varphi(\tau) \}$.

Examples: $h_k \approx \frac{1}{\sqrt{k+1}}$, $\Delta_T \leq O\left(\frac{1}{\sqrt{T}}\right)$.

Conclusion

Conclusion

1. Relative Smoothness Condition provides a new framework for development of new and efficient optimization schemes.

Conclusion

1. Relative Smoothness Condition provides a new framework for development of new and efficient optimization schemes.

2. It is able to fit well the properties of the particular objective function.

Conclusion

1. Relative Smoothness Condition provides a new framework for development of new and efficient optimization schemes.

2. It is able to fit well the properties of the particular objective function.

3. It is very new.

Conclusion

1. Relative Smoothness Condition provides a new framework for development of new and efficient optimization schemes.

2. It is able to fit well the properties of the particular objective function.

3. It is very new. We are looking for the new and motivating applications.

Conclusion

1. Relative Smoothness Condition provides a new framework for development of new and efficient optimization schemes.

2. It is able to fit well the properties of the particular objective function.

3. It is very new. We are looking for the new and motivating applications.

THANK YOU FOR YOUR ATTENTION!