Introductory Course on Non-smooth Optimisation

Lecture 03 - Krasnosel'skiĭ-Mann iteration

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

Table of contents

Recap

Recap of descent methods

- include gradient descent, proximal gradient descent.
- convergence (rate) properties
  - objective function value
    - $O(1/k)$ convergence rate.
    - optimal $O(1/k^2)$ convergence rate.
  - sequence
    - $O(1/\sqrt{k})$ convergence rate.
    - optimal $O(1/k)$ convergence rate.
  - linear convergence under *e.g.* strong convexity.

NB: end of happiness, most of the above results, especially for objective function values, will not be true for non-descent type methods.

Operator splitting

Consider the problem

$$\min_{x\in\mathbb{R}^n} \mu_1\|x\|_1 + \mu_2\|\nabla x\|_1 + \frac{1}{2}\|Ax - f\|^2.$$

In 1D, both

$$\mathrm{prox}_{\gamma\|\cdot\|_1}(\cdot) \quad \text{and} \quad \mathrm{prox}_{\gamma\|\nabla\cdot\|_1}(\cdot)$$

have close form solution. However, not for

$$\mathrm{prox}_{\gamma(\|\cdot\|_1+\|\nabla\cdot\|_1)}(\cdot).$$

Operator splitting design properly structured scheme such that

- the proximal mapping of non-smooth functions are evaluated separated.
- gradient descent is applied to the smooth part.

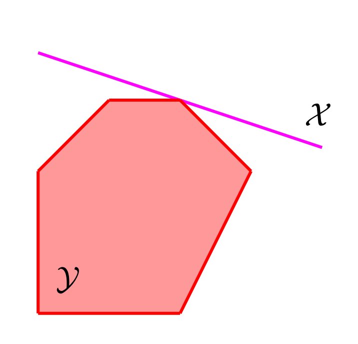
Outline

Feasibility problem

Consider finding a common point

$$\text{find} \;\; x \in \mathcal{X} \cap \mathcal{Y},$$

where $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^n$ are two closed and convex sets.



Equivalent formulation

$$\min_{x\in\mathbb{R}^n} \iota_{\mathcal{X}}(x) + \iota_{\mathcal{Y}}(x).$$

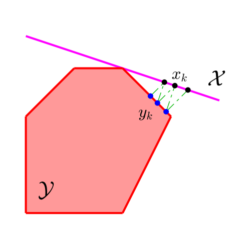
Method of alternating projection (MAP)

initial : $x_0 \in \mathcal{X}$;

repeat :

1. Projection onto $\mathcal{Y}$: $y_k = \mathcal{P}_{\mathcal{Y}}(x_k)$
2. Projection onto $\mathcal{X}$: $x_{k+1} = \mathcal{P}_{\mathcal{X}}(y_k)$

until : stopping criterion is satisfied.



- The projection onto two sets are computed separately.
- Stopping criterion: $\|x_k - x_{k-1}\| \le \epsilon$.

MAP

$$x_{k+1} = \mathcal{P}_{\mathcal{X}} \circ \mathcal{P}_{\mathcal{Y}}(x_k).$$

Convergence proerties

- convergence result for the objective function value?
- convergence of the sequences $\{x_k\}_{k\in\mathbb{N}}, \{y_k\}_{k\in\mathbb{N}}$?

Outline

Given two non-empty sets $\mathcal{X}, \mathcal{U} \subseteq \mathbb{R}^n$, $A : \mathcal{X} \rightrightarrows \mathcal{U}$ is called set-valued operator if A maps every point in $\mathcal{X}$ to a subset of $\mathcal{U}$, *i.e.*

$$A : \mathcal{X} \rightrightarrows \mathcal{U},\ x \in \mathcal{X} \mapsto A(x) \subseteq \mathcal{U}.$$

- The graph of A is defined by

$$\mathrm{gra}(A) \stackrel{\mathrm{def}}{=} \big\{(x, u) \in \mathcal{X} \times \mathcal{U} : u \in A(x)\big\}.$$

- The domain and range of A are

$$\mathrm{dom}(A) \stackrel{\mathrm{def}}{=} \big\{x \in \mathcal{X} : A(x) \neq \emptyset\big\},\ \mathrm{ran}(A) \stackrel{\mathrm{def}}{=} A(\mathcal{X}).$$

- The inverse of A defined through its graph

$$\mathrm{gra}(A^{-1}) \stackrel{\mathrm{def}}{=} \big\{(u, x) \in \mathcal{U} \times \mathcal{X} : u \in A(x)\big\}.$$

- The set of zeros of A are the points such that

$$\mathrm{zer}(A) \stackrel{\mathrm{def}}{=} A^{-1}(0) = \big\{x \in \mathcal{X} : 0 \in A(x)\big\}.$$

Monotone operator

Let $\mathcal{X}, \mathcal{U} \subseteq \mathbb{R}^n$ be two non-empty convex sets, $A : \mathcal{X} \rightrightarrows \mathcal{U}$ is monotone if

$$\langle x - y,\, u - v \rangle \geq 0,\ \ \forall (x, u), (y, v) \in \mathrm{gra}(A).$$

It is moreover maximal monotone if $\mathrm{gra}(A)$ is not strictly contained in the graph of any other monotone operators.

A is called α-strongly monotone for some $\kappa > 0$ if

$$\langle x - y,\, u - v \rangle \geq \kappa \|x - y\|^2.$$

Lemma

Let $R \in \Gamma_0$, then ∂R is maximal monotone.

Cocoercive operator

An operator $B : \mathbb{R}^n \to \mathbb{R}^n$ is called β-cocoercive if there exists $\beta > 0$ such that

$$\beta\|B(x) - B(y)\|^2 \le \langle B(x) - B(y),\, x - y\rangle, \quad \forall x, y \in \mathbb{R}^n.$$

The above equation implies that B is $(1/\beta)$-Lipschitz continuous.

Baillon-Haddad theorem

Let $F \in C_L^1$, then ∇F is β-cocoercive.

Lemma

Let $C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be β-strongly monotone, then its inverse C^{-1} is β-cocoercive.

Resolvent

Let $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone operator and $\gamma > 0$, the resolvent of A is defined by

$$\mathcal{J}_A \stackrel{\text{def}}{=} (\mathrm{Id} + A)^{-1}.$$

The reflection of $\mathcal{J}_A$ is defined by

$$\mathcal{R}_A \stackrel{\text{def}}{=} 2\mathcal{J}_A - \mathrm{Id}.$$

Given a function $R \in \Gamma_0$ and its sub-differential ∂R,

$$\mathrm{prox}_R = \mathcal{J}_{\partial R}.$$

Set of fixed points, $x = \mathrm{prox}_R(x)$

$$\mathrm{fix}(\mathrm{prox}_R) = \mathrm{fix}(\mathcal{J}_{\partial R}) = \mathrm{zer}(\partial R).$$

Yosida approximation

Let $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone operator and $\gamma > 0$, the Yosida approximation of A with γ is

$$ {}^{\gamma}A \stackrel{\text{def}}{=} \frac{1}{\gamma}(\mathrm{Id} - \mathcal{J}_{\gamma A}) = (\gamma \mathrm{Id} + A^{-1})^{-1} = \mathcal{J}_{A^{-1}/\gamma}(\cdot/\gamma). $$

Moreover,

$$ \mathrm{Id} = \mathcal{J}_{\gamma A}(\cdot) + \gamma \mathcal{J}_{A^{-1}/\gamma}\Big(\frac{\cdot}{\gamma}\Big). $$

- ${}^{\gamma}A$ is γ-cocoercive

Non-expansive operator

An operator $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ is called non-expansive if it is 1-Lipschitz continuous, *i.e.*

$$\|\mathcal{T}(x) - \mathcal{T}(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

For any $\alpha \in]0, 1[$, $\mathcal{T}$ is α-averaged if there exists a non-expansive operator $\mathcal{T}'$ such that

$$\mathcal{T} = \alpha\mathcal{T}' + (1 - \alpha)\mathrm{Id}.$$

- $\mathcal{A}(\alpha)$ denotes the class of α-averaged operators on $\mathbb{R}^n$.
- $\mathcal{A}(\frac{1}{2})$ is the class of firmly non-expansive operators.

Lemma

Let $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be non-expansive and $\alpha \in]0, 1[$. The following statements are equivalent:

- $\mathcal{T}$ is α-averaged non-expansive.
- The operator
$$\big(1 - \tfrac{1}{\alpha}\big)\mathrm{Id} + \tfrac{1}{\alpha}\mathcal{T}$$
is non-expansive.
- For any $x, y \in \mathbb{R}^n$,
$$\|\mathcal{T}(x) - \mathcal{T}(y)\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|(\mathrm{Id} - \mathcal{T})(x) - (\mathrm{Id} - \mathcal{T})(y)\|^2.$$

$\mathcal{A}(\alpha)$ is closed under relaxations, convex combinations and compositions.

Lemma

Let $m \in \mathbb{N}_+$, $\{\mathcal{T}_i\}_{i\in\{1,\ldots,m\}}$ be non-expansive operators on $\mathbb{R}^n$, $(\omega_i)_i \in]0,1]^m$ and $\sum_i \omega_i = 1$, and $(\alpha_i)_i \in]0,1]^m$ such that $\mathcal{T}_i \in \mathcal{A}(\alpha_i)$, $i \in \{1,...,m\}$. Then,

- $\mathrm{Id} + \lambda_i(\mathcal{T}_i - \mathrm{Id}) \in \mathcal{A}(\lambda_i\alpha_i),\ \lambda_i \in]0, \frac{1}{\alpha_i}[\quad \text{and} \quad i \in \{1,...,m\}$.
- $\sum_i \omega_i \mathcal{T}_i \in \mathcal{A}(\alpha)$ with $\alpha = \max_i \alpha_i$.
- $\mathcal{T}_1 \cdots \mathcal{T}_m \in \mathcal{A}(\alpha)$ with $\alpha = \frac{m}{m-1+1/\max_{i\in\{1,\ldots,m\}} \alpha_i}$.

Remark For the composition of two averaged operators, a sharper bound of α can be obtained,

$$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in]0, 1[.$$

Lemma

Let $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be non-expansive. The following statements are equivalent:

- $\mathcal{T}$ is firmly non-expansive.
- $\mathrm{Id} - \mathcal{T}$ is firmly non-expansive.
- $2\mathcal{T} - \mathrm{Id}$ is non-expansive.
- $\|\mathcal{T}(x) - \mathcal{T}(y)\|^2 \leq \langle \mathcal{T}(x) - \mathcal{T}(y),\, x - y\rangle, \forall x, y \in \mathbb{R}^n$.
- $\mathcal{T}$ is the resolvent of a maximal monotone operator A, *i.e.* $\mathcal{T} = \mathcal{J}_A$.

Lemma

Let operator $B : \mathbb{R}^n \to \mathbb{R}^n$ be β-cocoercive for some $\beta > 0$. Then

- $\beta B \in \mathcal{A}(\frac{1}{2})$, *i.e.* is firmly non-expansive.
- $\mathrm{Id} - \gamma B \in \mathcal{A}(\frac{\gamma}{2\beta})$ for $\gamma \in]0, 2\beta[$.

Outline

Fixed point

Let $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be a non-expansive operator, $x \in \mathbb{R}^n$ is called the fixed point of $\mathcal{T}$ if

$$x = \mathcal{T}(x).$$

The set of fixed points of $\mathcal{T}$ is denoted as $\mathrm{fix}(\mathcal{T})$.

- $\mathrm{fix}(\mathcal{T})$ may be empty, *e.g.* translation by a non-zero vector.

Lemma

Let $\mathcal{X}$ be a non-empty bounded closed convex subset of $\mathbb{R}^n$ and $\mathcal{T} : \mathcal{X} \to \mathbb{R}^n$ be a non-expansive operator, then $\mathrm{fix}(\mathcal{T}) \neq \emptyset$.

Lemma

Let $\mathcal{X}$ be a non-empty closed convex subset of $\mathbb{R}^n$ and $\mathcal{T} : \mathcal{X} \to \mathbb{R}^n$ be a non-expansive operator, then $\mathrm{fix}(\mathcal{T})$ is closed and convex.

Fixed-point iteration

$$x_{k+1} = \mathcal{T}(x_k).$$

However, it may not converge, *e.g.* $\mathcal{T} = -\mathrm{Id}$...

Krasnosel'skiĭ-Mann iteration

Let $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be a non-expansive operator such that $\mathrm{fix}(\mathcal{T}) \neq \emptyset$. Let $\lambda_k \in [0, 1]$ and choose x_0 arbitrarily from $\mathbb{R}^n$, then the Krasnosel'skiĭ-Mann iteration of $\mathcal{T}$ reads

$$x_{k+1} = x_k + \lambda_k(\mathcal{T}(x_k) - x_k).$$

- If $\mathcal{T} \in \mathcal{A}(\alpha)$, then $\lambda_k \in [0, 1/\alpha]$

Fejér monotonicity

Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a non-empty set and $\{x_k\}_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Then

- $\{x_k\}_{k\in\mathbb{N}}$ is Fejér monotone with respect to $\mathcal{S}$ if

$$\|x_{k+1} - x\| \leq \|x_k - x\|, \quad \forall x \in \mathcal{S}, \forall k \in \mathbb{N}.$$

- $\{x_k\}_{k\in\mathbb{N}}$ is quasi-Fejér monotone with respect to $\mathcal{S}$, if there exists a summable sequence $\{\epsilon_k\}_{k\in\mathbb{N}} \in \ell_+^1$ such that

$$\forall k \in \mathbb{N}, \quad \|x_{k+1} - x\| \leq \|x_k - x\| + \epsilon_k, \quad \forall x \in \mathcal{S}.$$

Example Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a non-empty convex set, and $\mathcal{T} : \mathcal{X} \to \mathcal{X}$ be a non-expansive operator such that $\mathrm{fix}(\mathcal{T}) \neq \emptyset$. The sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by

$$x_{k+1} = \mathcal{T}(x_k)$$

is Fejér monotone with respect to $\mathrm{fix}(\mathcal{T})$.

Lemma

Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a non-empty set and $\{x_k\}_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Assume the $\{x_k\}_{k\in\mathbb{N}}$ is quasi-Fejér monotone with respect to $\mathcal{S}$, then the following holds

- $\{x_k\}_{k\in\mathbb{N}}$ is bounded.
- $\|x_k - x\|$ is bounded for any $x \in \mathcal{S}$.
- $\{\mathrm{dist}(x_k, \mathcal{S})\}_{k\in\mathbb{N}}$ is decreasing and convergent.

If every sequential cluster point of $\{x_k\}_{k\in\mathbb{N}}$ belongs to $\mathcal{S}$, then $\{x_k\}_{k\in\mathbb{N}}$ converges to a point in $\mathcal{S}$.

- Weak convergence in general real Hilbert space

Convergence

Let $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be a non-expansive operator such that $\mathrm{fix}(\mathcal{T}) \neq \emptyset$. Consider the Krasnosel'skiĭ-Mann iteration of $\mathcal{T}$, and choose $\lambda_k \in [0, 1]$ such that

$$\sum_{k \in \mathbb{N}} \lambda_k (1 - \lambda_k) = +\infty,$$

then the following holds

- $\{x_k\}_{k \in \mathbb{N}}$ is Fejér monotone with respect to $\mathrm{fix}(\mathcal{T})$.
- $\{x_k - \mathcal{T}(x_k)\}_{k \in \mathbb{N}}$ converges strongly to 0.
- $\{x_k\}_{k \in \mathbb{N}}$ converges to a point in $\mathrm{fix}(\mathcal{T})$.

Remark When $\mathcal{T}$ is α-averaged, then

$$\lambda_k \in [0, 1/\alpha] \ \text{ such that } \ \textstyle\sum_{k \in \mathbb{N}} \lambda_k (1/\alpha - \lambda_k) = +\infty.$$

- Krasnosel'skiĭ-Mann iteration with constant relaxation

$$\begin{aligned} x_{k+1} &= x_k + \lambda\big(\mathcal{T}(x_k) - x_k\big) \\ &= \big((1-\lambda)\mathrm{Id} + \lambda\mathcal{T}\big)(x_k). \end{aligned}$$

- Denote $\mathcal{T}_\lambda = (1-\lambda)\mathrm{Id} + \lambda\mathcal{T}$, and define residual

$$e_k = (\mathrm{Id} - \mathcal{T})(x_k) = \frac{1}{\lambda}(x_k - x_{k+1}).$$

  - $\mathcal{T}_\lambda \in \mathcal{A}(\lambda)$ if $\lambda \in]0,1[$. If $\mathcal{T} \in \mathcal{A}(\alpha)$, then $\mathcal{T}_\lambda \in \mathcal{A}(\lambda\alpha)$.
  - For any $x^\star \in \mathrm{fix}(\mathcal{T})$,

$$x^\star \in \mathrm{fix}(\mathcal{T}) \Leftrightarrow x^\star \in \mathrm{fix}(\mathcal{T}_\lambda) \Leftrightarrow x^\star \in \mathrm{zer}(\mathrm{Id} - \mathcal{T}).$$

  - If $\lambda \in [\epsilon, 1-\epsilon]$, $\epsilon \in]0, 1/2]$,
    - e_k converges to 0.
    - $\{x_k\}_{k\in\mathbb{N}}$ is quasi-Fejér monotone with respect to $\mathrm{fix}(\mathcal{T})$, and converges to a point $x^\star \in \mathrm{fix}(\mathcal{T})$.

Pointwise convergence rate

Rate of $\|e_k\|^2$:

- For residual

$$\|e_{k+1}\|^2 \leq \|e_k\|^2 - \frac{1-\lambda}{\lambda}\|e_k - e_{k+1}\|^2.$$

- $\mathcal{T}_\lambda \in \mathcal{A}(\lambda)$, $\tau = \lambda(1-\lambda)$

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - \tau\|e_k\|^2.$$

- Summation

$$(k+1)\|e_k\|^2 \leq \tau \sum_{i=0}^{k} \|e_i\|^2 \leq \|x_0 - x^\star\|^2 - \|x_{k+1} - x^\star\|^2.$$

- Rate

$$\|e_k\|^2 \leq \frac{\|x_0 - x^\star\|^2}{k+1}.$$

NB: if $T \in \mathcal{A}(\alpha)$, then the above holds for $\lambda \in [\epsilon, 1/\alpha - \epsilon]$.

Ergodic convergence rate

Define $\bar{e}_k = \frac{1}{k+1}\sum_{i=0}^{k} e_i$.

- Boundedness

$$\begin{aligned}\|x_{k+1} - x^\star\| = \|\mathcal{T}_\lambda(x_k) - \mathcal{T}_\lambda(x^\star)\| &\leq \|x_k - x^\star\| \\ &\leq \|x_0 - x^\star\|.\end{aligned}$$

- $\lambda e_k = x_k - x_{k+1}$

$$\begin{aligned}\|\bar{e}_k\| = \frac{1}{k+1}\|\textstyle\sum_{i=0}^{k} e_i\| &= \frac{1}{\lambda(k+1)}\|\textstyle\sum_{i=0}^{k}(x_i - x_{i+1})\| \\ &= \frac{1}{\lambda(k+1)}\|x_0 - x_{k+1}\| \\ &\leq \frac{1}{\lambda(k+1)}(\|x_0 - x^\star\| + \|x_{k+1} - x^\star\|) \\ &\leq \frac{2\|x_0 - x^\star\|}{\lambda(k+1)}.\end{aligned}$$

NB: both rates (pointwise and ergodic) can be extended to the inexact case...

Metric sub-regularity

A set-valued mapping $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called metrically sub-regular at $\bar{x}$ for $\bar{u} \in A(\bar{x})$ if there exists $\kappa \geq 0$ along with neighbourhood $\mathcal{X}$ of $\bar{x}$ such that

$$\mathrm{dist}(x, A^{-1}(\bar{u})) \leq \kappa \, \mathrm{dist}(\bar{u}, A(x)), \ \ \forall x \in \mathcal{X}.$$

The infimum of all κ such that above holds is called the modulus of metric sub-regularity, and denoted by $\mathrm{subreg}(A; \bar{x}|\bar{u})$.

Example Let $F \in S^1_{\alpha,L}$ and $A = \gamma\nabla F$ with $\gamma \leq 1/L$: $\bar{x} = \mathrm{argmin}_{\mathbb{R}^n} F$ and $\bar{u} = 0$,

$$\begin{aligned}\mathrm{dist}(\bar{u}, A(x)) &= \|\gamma\nabla F(x) - \gamma\nabla F(\bar{x})\| \\ &\geq \gamma\alpha\|x - y\|\end{aligned}$$

Let $x^\star \in \mathrm{fix}(\mathcal{T})$, suppose $\mathcal{T}' \stackrel{\mathrm{def}}{=} \mathrm{Id} - \mathcal{T}$ is metrically sub-regular at $x^\star$ with neighbourhood $\mathcal{X}$ of $x^\star$, let $\kappa > \mathrm{subreg}(\mathcal{T}'; x^\star|0)$:

- $0 = \mathcal{T}'(x^\star)$, $\mathcal{T}'^{-1}(0) = \mathrm{fix}(\mathcal{T})$

$$\mathrm{dist}(x, \mathrm{fix}(\mathcal{T})) \leq \kappa \mathrm{dist}(0, \mathcal{T}'(x)) = \kappa\|x - \mathcal{T}(x)\|.$$

- Denote $d_k = \mathrm{dist}(x_k, \mathrm{fix}(\mathcal{T}))$, $\bar{x} \in \mathrm{fix}(\mathcal{T})$ such that $d_k = \|x_{k+1} - \bar{x}\|$,

$$\begin{aligned} d_{k+1}^2 \leq \|x_{k+1} - \bar{x}\|^2 &\leq \|x_k - \bar{x}\|^2 - \tau\|\mathcal{T}'(x_k) - \mathcal{T}'(\bar{x})\|^2 \\ &\leq d_k^2 - \tfrac{\tau}{\kappa^2} d_k^2 \\ &= \big(1 - \tfrac{\tau}{\kappa^2}\big) d_k^2. \end{aligned}$$

NB: As metric sub-regularity is a local propery, the linear convergence will happen only when x_k is close enough to $\mathrm{fix}(\mathcal{T})$.

Optimal relaxation parameter?

Consider $\lambda_k \in [0,1]$ and $x_{k+1} = (1-\lambda_k)x_k + \lambda_k \mathcal{T}(x_k)$. Then

$$\begin{aligned} \|x_{k+1} - x^\star\|^2 &= \|(1-\lambda_k)(x_k - x^\star) + \lambda_k(\mathcal{T}(x_k) - x^\star)\|^2 \\ &= (1-\lambda_k)\|x_k - x^\star\|^2 + \lambda_k\|\mathcal{T}(x_k) - x^\star\|^2 \\ &\qquad - \lambda_k(1-\lambda_k)\|x_k - \mathcal{T}(x_k)\|^2 \\ &= \lambda_k^2\|x_k - \mathcal{T}(x_k)\|^2 \\ &\qquad - \lambda_k\big(\|x_k - x^\star\|^2 - \|\mathcal{T}(x_k) - x^\star\|^2 + \|x_k - \mathcal{T}(x_k)\|^2\big) + \|x_k - x^\star\|^2 \end{aligned}$$

which is a quadratic fucntion of λ_k, and minimises at

$$\lambda = \frac{1}{2} + \frac{\|x_k - x^\star\|^2 - \|\mathcal{T}(x_k) - x^\star\|^2}{2\|x_k - \mathcal{T}(x_k)\|^2}.$$

Approximation:

$$\lambda = \frac{1}{2} + \frac{\|x_k - \mathcal{T}(x_k)\|^2 - \|\mathcal{T}(x_k) - \mathcal{T}^2(x_k)\|^2}{2\|(x_k - \mathcal{T}(x_k)) - (\mathcal{T}(x_k) - \mathcal{T}^2(x_k))\|^2}.$$

Outline

An inertial Krasnosel'skiĭ-Mann iteration

Initial : $x_0 \in \mathbb{R}^n, x_{-1} = x_0$;

$$\begin{aligned} y_k &= x_k + a_k(x_k - x_{k-1}),\ a_k \in [0,1], \\ z_k &= x_k + b_k(x_k - x_{k-1}),\ b_k \in [0,1], \\ x_{k+1} &= (1-\lambda_k)y_k + \lambda_k \mathcal{T}(z_k),\ \lambda_k \in [0,1]. \end{aligned}$$

- Covers heavy-ball method, Nesterov's scheme, inertial FB and FISTA.
- Convergence analysis is much harder than the inertial version of descent methods.
- No convergence rate.
- May perform very poorly in practice, slower than the original scheme.

A multi-step inertial Krasnosel'skiĭ-Mann iteration

Initial : $x_0 \in \mathbb{R}^n, x_{-1} = x_0$ and $\gamma \in]0, 2/L[$;

$$
\begin{aligned}
y_k &= x_k + a_{0,k}(x_k - x_{k-1}) + a_{1,k}(x_{k-1} - x_{k-2}) + \cdots\ , \\
z_k &= x_k + b_{0,k}(x_k - x_{k-1}) + b_{1,k}(x_{k-1} - x_{k-2}) + \cdots\ , \\
x_{k+1} &= (1 - \lambda_k)y_k + \lambda_k \mathcal{T}(z_k),\ \lambda_k \in [0, 1].
\end{aligned}
$$

- Even harder to analyse convergence.
- No rate.
- However, can outperform the original scheme...

Convergence

- Conditional convergence, $i = 0, 1, \ldots$

$$\sum_{k\in\mathbb{N}} \max\left\{ \max_i |a_{i,k}|, \max_i |b_{i,k}| \right\} \sum_i \|x_{k-i} - x_{k-i-1}\| < +\infty.$$

- Online updating rule

$$a_{i,k} = \min\left\{a_i, c_{i,k}\right\}$$

where

$$c_{i,k} = \frac{c_i}{k^{1+\delta} \sum_i \|x_{k-i} - x_{k-i-1}\|}, \ \delta > 0.$$

Reference

- H. Bauschke and P. L. Combettes. “Convex Analysis and Monotone Operator Theory in Hilbert Spaces”. Springer, 2011.
- P. L. Lions and B. Mercier. “Splitting algorithms for the sum of two nonlinear operators”. SIAM Journal on Numerical Analysis, 16(6):964–979, 1979.
- J. Liang, J. Fadili, and G. Peyré. “Convergence rates with inexact non-expansive operators”. Mathematical Programming 159.1-2 (2016): 403-434.