November 29, 2022

Johann Radon Institute (RICAM)

Fixed point strategies for nonconvexly regularized sparse estimation and hierarchical convex optimization

Isao Yamada

Tokyo Institute of Technology

Many thanks : This talk is designed mainly based on recent joint works with excellent collaborators in Yamada laboratory of Tokyo Tech !

Contents of this talk

1. Why fixed point expressions of closed convex sets ?

2. A fixed point strategy for
   sparsity-aware estimation with nonconvex regularizations

3. A fixed point strategy for
   Hierarchical Convex Optimization

$\mathcal{H}$: Real Hilbert space

A mapping $T : \mathcal{H} \to \mathcal{H}$ is called a **nonexpansive operator** if

$$(\forall x, y \in \mathcal{H}) \ \|T(x) - T(y)\| \leq \|x - y\|.$$

If $(\exists \kappa \in (0,1), \ \forall x, y \in \mathcal{H}) \ \|T(x) - T(y)\| \leq \kappa \|x - y\|$, T is called a **contraction operator**

and guaranteed to have a unique fixed point in $\mathcal{H}$ (Banach-Picard) !

If a **nonexpansive operator** has a nonempty fixed point set

$$\mathrm{Fix}(T) := \{z \in \mathcal{H} \mid T(z) = z\} \neq \emptyset,$$

$\Rightarrow$ $\mathrm{Fix}(T)$ becomes a **closed convex** subset of $\mathcal{H}$!

Any case can happen for general Nonexpansive operators

Case 1: T **has no fixed point in** $\mathcal{H}$**.**

Case 2: T **has a unique fixed point in** $\mathcal{H}$**.**

Case 3: T **has two distinct fixed points in** $\mathcal{H}$**,**
which implies $\sharp(\mathrm{Fix}(T)) \geq \aleph$ **!**

Fact (Intersection of fixed point sets of nonexpansive operators)

Let $T_i (i=1,2)$ be α_i -averaged nonexpansive operators, i.e.,

$$\left.\begin{array}{l} T_1 = (1-\alpha_1)\mathrm{Id} + \alpha_1 R_1 \ \ (\exists R_1\text{: Nonexpansive}, \exists \alpha_1 \in (0,1)) \\ T_2 = (1-\alpha_2)\mathrm{Id} + \alpha_2 R_2 \ \ (\exists R_2\text{: Nonexpansive}, \exists \alpha_2 \in (0,1)) \end{array}\right\}, \textbf{ and } \mathrm{Fix}(T_1)\cap\mathrm{Fix}(T_2) \neq \varnothing$$

$$\Rightarrow \mathrm{Fix}(T_1)\cap\mathrm{Fix}(T_2) = \mathrm{Fix}(T_1\circ T_2) = \mathrm{Fix}\left((1-\omega)T_1 + \omega T_2\right) \quad (\forall \omega \in (0,1))$$

NOTE: Any nonexpansive operator can be modified as α_i-averaged nonexpansive operator without changing its fixed point set via convex combination with Id !

⇒ **A computable nonexpansive operator** can serve as **a mathematically precise language** to express closed convex sets !

Proposition (Composition of averaged operators) [Ogura-Yamada 2002] [Combettes-Yamada 2015]

Let $\alpha_1, \alpha_2 \in (0,1)$ and let $T_i : \mathcal{H} \to \mathcal{H}$ $(i=1,2)$ be α_i-averaged.

$$\Rightarrow \begin{cases} T_1 \circ T_2 \text{ is } \frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}\text{-averaged.} \\ (\forall \omega \in (0,1)) \quad (1-\omega)T_1 + \omega T_2 \text{ is } ((1-\omega)\alpha_1 + \omega\alpha_2)\text{-averaged.} \end{cases}$$

(A) How can we compute a valuable fixed point in $\mathrm{Fix}(T) := \{z \in \mathcal{H} \mid T(z) = z\}$?

Fact (Krasnosel'skiĭ('55)-Mann('53))

Let $T : \mathcal{H} \to \mathcal{H}$ be nonexpansive with $\mathrm{Fix}(T) \neq \varnothing$.

Then for any $x_0 \in \mathcal{H}$ and $\alpha \in (0,1)$, the sequence $(x_n)_{n=0}^{\infty}$ generated by

$$x_{n+1} = (1-\alpha)x_n + \alpha T(x_n)$$

converges weakly to a certain point in $\mathrm{Fix}(T)$.

Note : If T can be expressed as $T = (1-\beta)\mathrm{Id} + \beta R$ with some nonexpansive operator and $\beta \in (0,1)$,

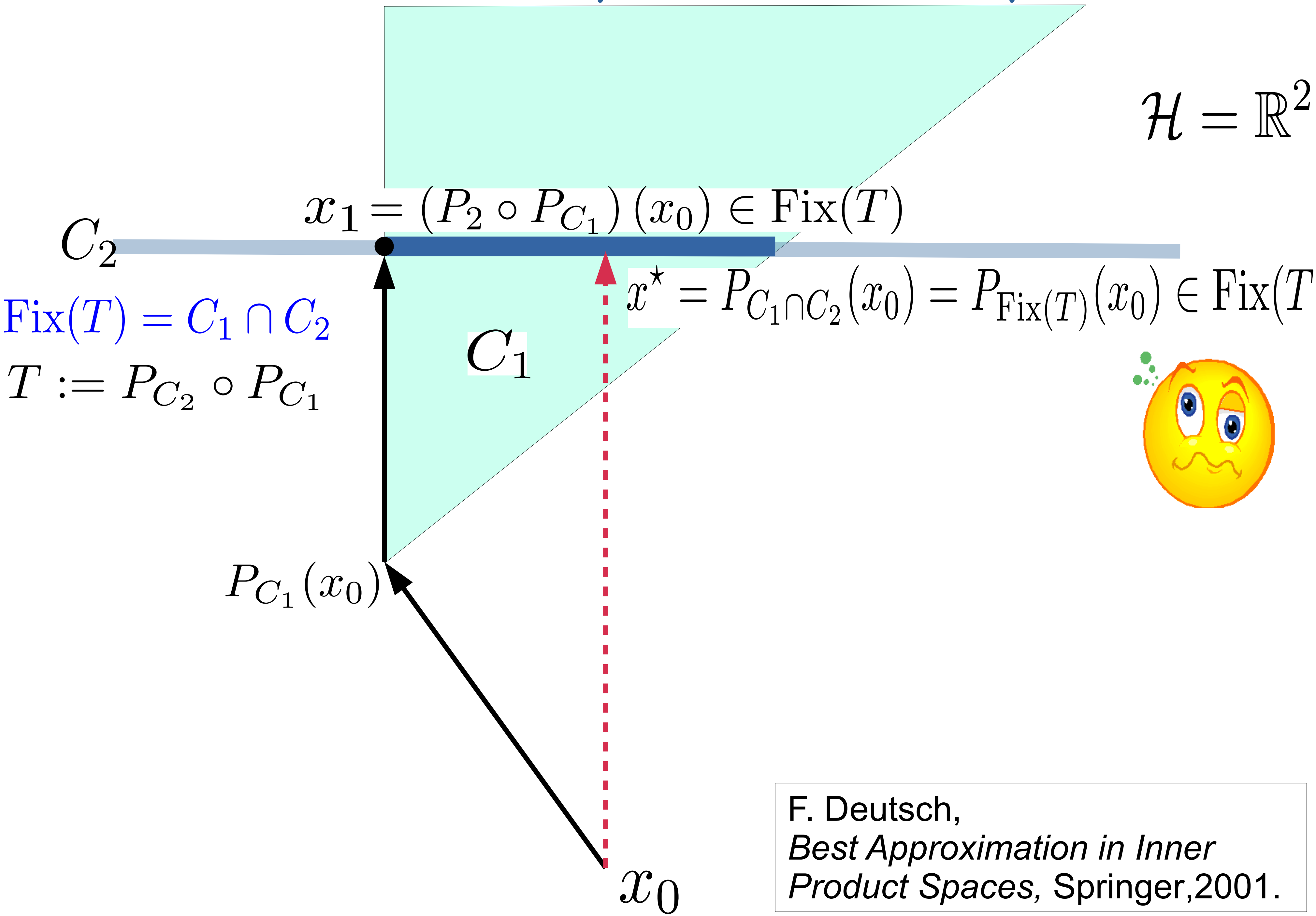
$$T^{n+1}(x_0) = (1-\beta)x_n + \beta R(x_n) \rightharpoonup \exists z \in \mathrm{Fix}(R) = \mathrm{Fix}(T)$$

(i) K-M iteration is a remarkable extensuin of Picard iteration !

(ii) POCS [Bregman 65] can be seen as an application of K-M iteration !

Remark : In general, the weak limit of K-M iteration does not enjoy any best approximation property among all fixed points in Fix(T) !

POCS (an instance of K-M iteration) does not converge to the nearest fixed point from the initial point !



F. Deutsch,
Best Approximation in Inner Product Spaces, Springer,2001.

(B) How can we design a computable nonexpansive operator whose fixed point set $\mathrm{Fix}(T)$ has great application value ?

Convex Optimization Problem

defined on a Real Hilbert Space $\mathcal{X}$

$$\text{Minimize } \varphi : \mathcal{X} \to (-\infty, \infty]$$

where $\varphi \in \Gamma_0(\mathcal{X})$,

i.e.,

Proper	$\mathrm{dom}(\varphi) := \{x \in \mathcal{X} \mid \varphi(x) < \infty\} \neq \varnothing$
Lower Semi-continuous	$(\forall \alpha \in \mathbb{R})\ \mathrm{lev}_{\leq \alpha}(\varphi) := \{x \in \mathcal{X} \mid \varphi(x) \leq \alpha\}$ is Closed in $\mathcal{X}$
Convex **Bowl-shaped function**	$(\forall x, y \in \mathrm{dom}(\varphi), \forall \lambda \in (0,1))$ $\varphi(\lambda x + (1-\lambda)y) \leq \lambda \varphi(x) + (1-\lambda)\varphi(y)$

Convex optimization often found in signal processing and inverse problems

$$(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}}, \|\cdot\|_{\mathcal{X}}), (\mathcal{K}, \langle\cdot,\cdot\rangle_{\mathcal{K}}, \|\cdot\|_{\mathcal{K}}): \text{ Real Hilbert Spaces}$$

$$f \in \Gamma_0(\mathcal{X}),\ g \in \Gamma_0(\mathcal{K}),\ A : \mathcal{X} \to \mathcal{K}: \text{ Bdd linear}$$

$$\textbf{(P)} \quad \underset{x\in\mathcal{X}}{\text{minimize}}\ f(x) + g\circ A(x)$$

convex — convex

has been playing central roles in various esimation problems

because seemingly a much more general model

$$f \in \Gamma_0(\mathcal{X}),\ g_i \in \Gamma_0(\mathcal{K}_i),\ A_i : \mathcal{X} \to \mathcal{K}_i: \text{ Bdd linear}$$

$$\textbf{(Q)} \quad \underset{x\in\mathcal{X}}{\text{minimize}}\ f(x) + \sum_{i=1}^{M} g_i(A_i x)$$

can also be handled as an instance of (P) by

$$\mathcal{K} := \mathcal{K}_1 \times \cdots \times \mathcal{K}_M,\ g := \bigoplus_{i=1}^{M} g_i \text{ and } Ax := (A_1 x, \ldots, A_M x)$$

New Hilbert space — **Product Space**

New convex function — **Separable sum**

New linear operator

Building block to design computable nonexpansive operator

Let $f \in \Gamma_0(\mathcal{X})$, i.e., (Proper, lower-semicontinuous,) convex function on $\mathcal{X}$.

Proximity operator [Moreau 1962]

- The proximity operator : **(Innovative Generalization of Projection !)**

$$\mathrm{Prox}_f : \mathcal{X} \to \mathcal{X} : x \mapsto \underset{y \in \mathcal{X}}{\mathrm{argmin}} \left(f(y) + \frac{1}{2}\|x - y\|^2 \right).$$

is $\frac{1}{2}$-averaged nonexpansive, and satisfies

$$\mathrm{Fix}(\mathrm{Prox}_f) := \{x \in \mathcal{X} \mid \mathrm{Prox}_f(x) = x\} = \mathrm{Argmin}\, f$$

Good News

Closed form expressions of prox_f are available for many $f \in \Gamma_0(\mathcal{X})$, called *Proximable* (or *Prox-friendly*) functions.

See, e.g.,

G. CHIERCHIA, E. CHOUZENOUX, P. L. COMBETTES, J-C. PESQUET

http://proximity-operator.net/

Example (Indicator function ι_C)

For a nonempty closed convex set $C(\subset \mathcal{X})$,

the **proximity operator of the indicator function** :

$$\iota_C(x) := \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise} \end{cases}$$

is given by **projection** onto C, i.e.,

$$\mathrm{Prox}_{\iota_C} : \mathcal{X} \to \mathcal{X} : x \mapsto \operatorname*{argmin}_{y \in \mathcal{X}} \left(\iota_C(y) + \frac{1}{2}\|y - x\|^2 \right) = \operatorname*{argmin}_{y \in C} \frac{1}{2}\|y - x\|^2 =: P_C(x).$$

Example (ℓ_1-norm: A largest convex minorant of $\|\cdot\|_0$)

$$\|\cdot\|_1 : \mathbb{R}^n \to \mathbb{R}_+ : \mathbf{x} := (x_1, \ldots, x_n)^T \mapsto \sum_{i=1}^{n} |x_i|$$

$$\left[\mathrm{Prox}_{\gamma\|\cdot\|_1}(\mathbf{x})\right]_i = \mathrm{sgn}(x_i) \max\{|x_i| - \gamma, 0\}$$

soft-thresholding

$\|\cdot\|_1$ is Prox-friendly

Proximal Splitting

A key idea to design computable nonexpansive operators

1. In general, even if $f \in \Gamma_0(\mathcal{X})$ and $g \in \Gamma_0(\mathcal{K})$ are prox-friendly,
   and $A : \mathcal{X} \to \mathcal{K}$ is computable Bdd linear operator,

   $g \circ A \in \Gamma_0(\mathcal{X})$ and $f + g \circ A \in \Gamma_0(\mathcal{X})$ are not necessarily prox-friendly !

2. **Proximal Splitting** is the art of computational techniques
   mainly for expressions of the solution set of

$$\text{minimize } f + g \circ A \in \Gamma_0(\mathcal{X})$$

in terms of the **fixed point set of a computable nonexpansive operator**

with building blocks $\mathrm{Prox}_f : \mathcal{X} \to \mathcal{X}$ and $\mathrm{Prox}_g : \mathcal{K} \to \mathcal{K}$

See, e.g., H.H. Bauschke and P. L. Combettes. Convex Analysis and Monotone Operator Theory in Hilbert Spaces. 2nd ed., Springer, 2017.

Contents of this talk

1. Why fixed point expressions of closed convex sets ?

2. A fixed point strategy for
   sparsity-aware estimation with nonconvex regularizations

3. A fixed point strategy for
   Hierarchical Convex Optimization

Many tasks in sparsity-aware signal processing and inverse problems have been formulated as

sparsity-aware regularized least squares models

$$\underset{x\in\mathcal{X}}{\text{minimize}}\ J_{\Psi\circ\mathfrak{L}}(x) := \frac{1}{2}\|y - Ax\|_{\mathcal{Y}}^2 + \mu\Psi\circ\mathfrak{L}(x),\ \ \mu > 0, \qquad (1)$$

convex   **convex or nonconvex**

where $\mathcal{X},\mathcal{Y},\mathcal{Z}$: finite dimensional real Hilbert spaces, $y\in\mathcal{Y}$, $A\in\mathcal{B}(\mathcal{X},\mathcal{Y})$ (i.e. A is a bounded linear operator from $\mathcal{X}$ to $\mathcal{Y}$), $\mathfrak{L}\in\mathcal{B}(\mathcal{X},\mathcal{Z})$ and $\Psi:\mathcal{Z}\to\mathbb{R}_+$ is a certain **approximation of** $\|\cdot\|_0$ ($\sharp$ of nonzero entries)

Convexly regularized least squares
$\|\cdot\|_1$ is used as Ψ

Lasso [Tibshirani `96] ,
TV [ROF`92],
[Daubechies et al `04] …

This talk
Nonconvex regularization via Moreau enhancement

MCP [Zhang `10], GMC [Selesnick `17]
LiGME [AYY `20] …

GME matrix design

[CYY `22]

cLiGME model

[YYY `22]

$\Psi(\mathbf{x}) = \|(x_1, \ldots, x_n)\|_0 :=$ Number of nonzero-entries of $(x_1, \ldots, x_n) \in \mathbb{R}^n$

$$\underset{x \in \mathcal{X}}{\text{minimize}} \; J_{\|\cdot\|_0 \circ \mathfrak{L}}(x) := \frac{1}{2}\|y - Ax\|_{\mathcal{Y}}^2 + \mu\|\cdot\|_0 \circ \mathfrak{L}(x)$$

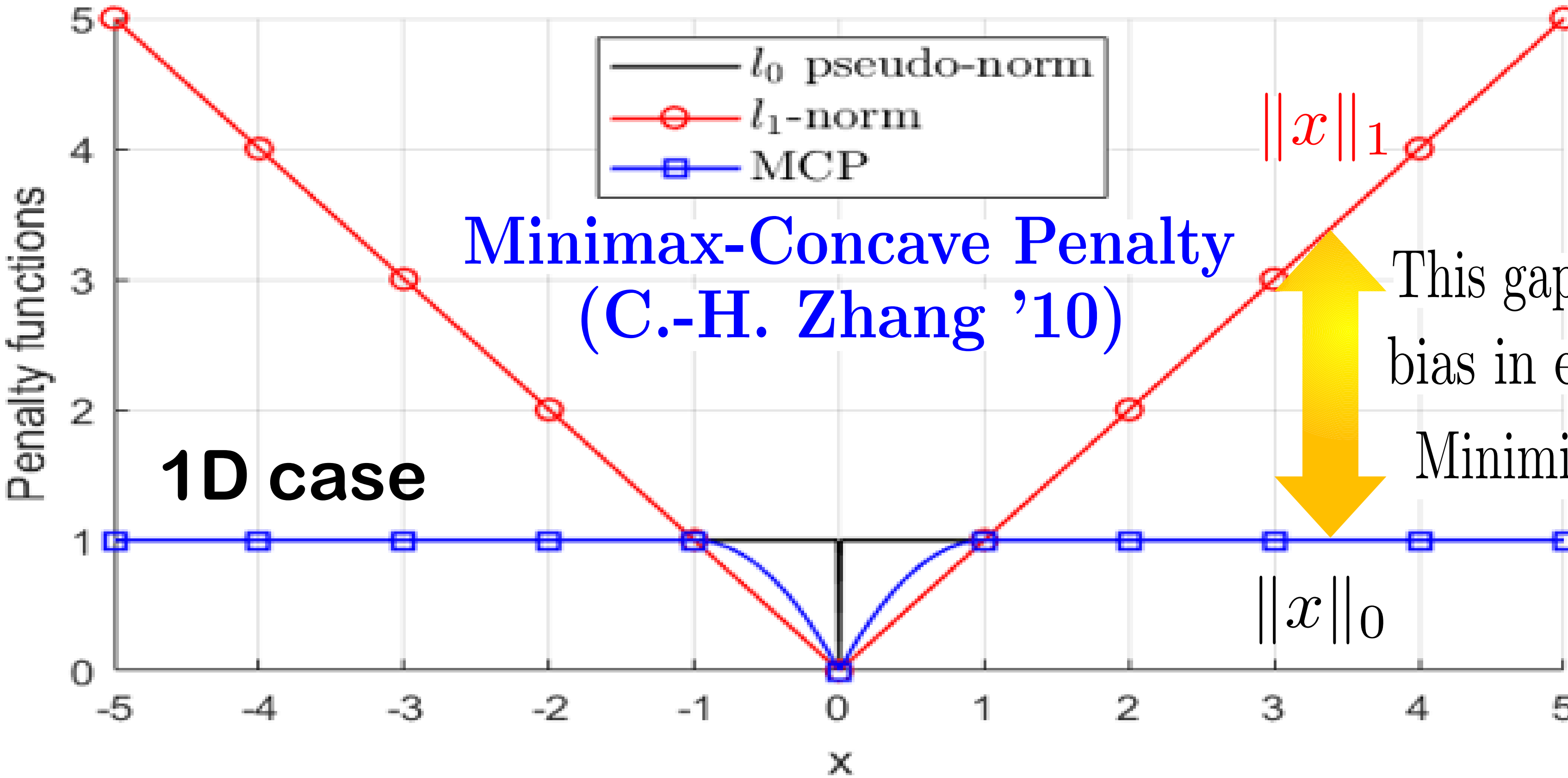
$\|\cdot\|_0$ is ideal criterion but this model is NP-hard

Largest convex minorant of $\|(x_1, \ldots, x_n)\|_0$ on $[-1, 1]^n$

$$\Psi(\mathbf{x}) = \|(x_1, \ldots, x_n)\|_1 := \sum_{i=1}^{n} |x_i|$$

Moreau enhancement (nonconvex)

$$\Psi(\mathbf{x}) = {}^{\gamma}(\|(x_1, \ldots, x_n)\|_1)_{\mathrm{MC}} := \sum_{i=1}^{n} {}^{\gamma}|x_i|_{\mathrm{MC}}$$



This gap causes large bias in estimation by Minimize $J_{\|\cdot\|_1 \circ \mathfrak{L}}(x)$

Minimax-Concave (MC) penalty [C.-H. Zhang 2010]

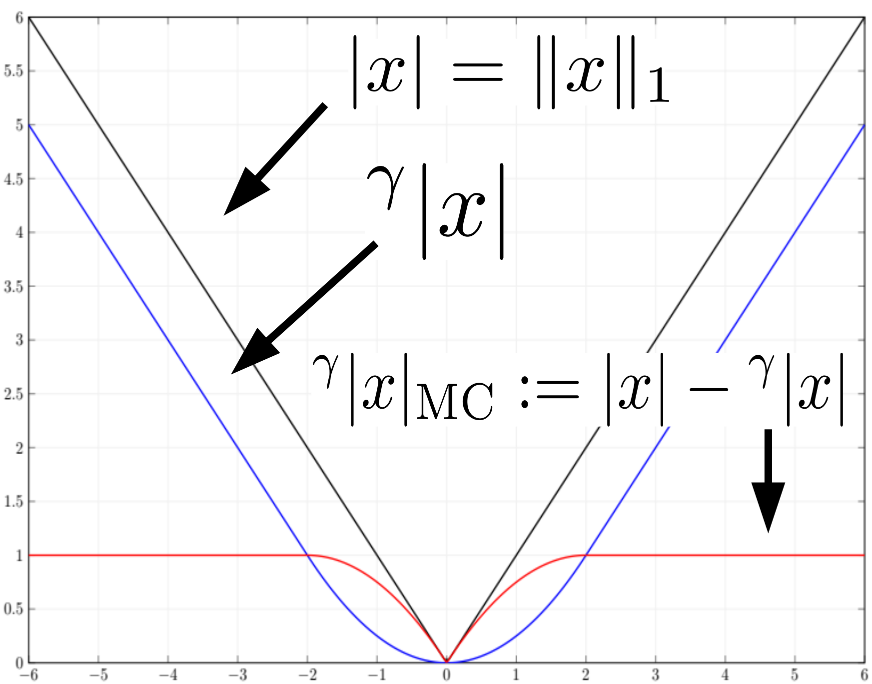
is a simplest 1D example of LiGME function $(\Psi = |\cdot| \text{ and } \mathfrak{L} = \mathrm{Id})$

Moreau envelope of $|\cdot|$

P.J.Huber, Ann. Math. Statist.'64

$$ {}^{\gamma}|x| := \min_{v \in \mathbb{R}} \left[|v| + \frac{1}{2\gamma}|x-v|^2 \right] = \begin{cases} \frac{1}{2\gamma}x^2, & \text{if } |x| \leq \gamma\ ; \\ |x| - \frac{1}{2}\gamma, & \text{otherwise.} \end{cases} $$

converges pointwise to $|\cdot|$ as $\gamma \downarrow 0$.



C.-H.Zhang, Ann. Statist.'10

Minimax-Concave penalty

$$ {}^{\gamma}|x|_{\mathrm{MC}} := |x| - {}^{\gamma}|x| $$
$$ = \begin{cases} |x| - \frac{1}{2\gamma}x^2, & \text{if } |x| \leq \gamma\ ; \\ \frac{1}{2}\gamma, & \text{otherwise.} \end{cases} $$

has been proposed as a nearly unbiased **nonconvex enhancement of the best convex sparsity promoting regularizer** ℓ_1-norm $\|\cdot\|_1$

LiGME is a Unified + Linearly involved extension

[Abe-Yamagishi-IY (Inverse Problems '20)]

For $\mathcal{X}, \mathcal{Z}, \widetilde{\mathcal{Z}}$: Hilbert spaces and $\Psi \in \Gamma_0(\mathcal{Z})$ [prox-friendly, coercive, $\operatorname{dom} \Psi = \mathcal{Z}$],

$$\left(B \in \mathcal{B}(\mathcal{Z}, \widetilde{\mathcal{Z}}),\ \mathfrak{L} \in \mathcal{B}(\mathcal{X}, \mathcal{Z})\right) \qquad \Psi_B \circ \mathfrak{L} : \mathcal{X} \to \mathbb{R} \qquad \textbf{(LiGME)}$$

where

$$\Psi_B(\cdot) := \Psi(\cdot) - \min_{v \in \mathcal{Z}} \left[\Psi(v) + \frac{1}{2} \| B(\cdot - v) \|^2 \right].$$

Generalized Moreau envelope of $\Psi(\cdot)$

Generalized Minimax-Concave (GMC) Penalty

[I. Selesnick *IEEE T-SP,* 2017]

$$(B \in \mathbb{R}^{m \times l}) \qquad (\| \cdot \|_1)_B(\mathbf{z}) := \|\mathbf{z}\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \| B(\mathbf{z} - \mathbf{v}) \|^2 \right] \qquad \textbf{(GMC)}$$

Minimax-Concave (MC) Penalty [C.-H. Zhang, Ann. Statist.'10]

$$(\gamma \in \mathbb{R}_{++}) \qquad {}^{\gamma}|z|_{\mathrm{MC}} : \mathbb{R} \to \mathbb{R} :\ z \mapsto |z| - \min_{v \in \mathbb{R}} \left[|v| + \frac{1}{2\gamma} |z - v|^2 \right] \qquad \textbf{(MC)}$$

LiGME model [J.Abe, M.Yamagishi, IY, Inverse Problems 2020]

$$\underset{x\in\mathcal{X}}{\text{minimize}}\ J_{\Psi_B\circ\mathfrak{L}}(x) := \frac{1}{2}\|y - Ax\|^2 + \mu\Psi_B \circ \mathfrak{L}(x),\ \ \mu > 0, \tag{2}$$

convex — nonconvex

where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \widetilde{\mathcal{Z}}$: Hilbert spaces, $y \in \mathcal{Y}$, $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, $\mathfrak{L} \in \mathcal{B}(\mathcal{X}, \mathcal{Z})$ and

$$\Psi_B(\cdot) := \Psi(\cdot) - \min_{v\in\mathcal{Z}}\left[\Psi(v) + \frac{1}{2}\|B(\cdot - v)\|^2\right],$$

Nonconvex

with $\Psi \in \Gamma_0(\mathcal{Z})$ [prox-friendly, coercive, $\operatorname{dom}\Psi = \mathcal{Z}$] and $B \in \mathcal{B}(\mathcal{Z}, \widetilde{\mathcal{Z}})$.

(B is a tuning parameter for Linearly involved Generalized Moreau Enhancement of Ψ)

Good News 1

With proper choice of $B \in \mathcal{B}(\mathcal{Z}, \widetilde{\mathcal{Z}})$, the **desired overall convexity can be achieved !**

Overall Convexity Condition for (2)

$A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L} \succeq O \Rightarrow J_{\Psi_B\circ\mathfrak{L}} \in \Gamma_0(\mathcal{X}) \Rightarrow$ Existence of minimizer of (2) is guaranteed under mild condition

In particular, if $\Psi \in \Gamma_0(\mathcal{Z})$ satisfies the condition as a norm of vector space $\mathcal{Z}$,

$A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L} \succeq 0 \Leftrightarrow J_{\Psi_B\circ\mathfrak{L}} \in \Gamma_0(\mathcal{X})$ [Abe, Yamagishi, IY (Inverse Problems 2020)]

Good News 2

$$\underset{x\in\mathcal{X}}{\text{minimize}}\ J_{\Psi_B\circ\mathfrak{L}}(x) := \frac{1}{2}\|y - Ax\|^2 + \mu\Psi_B\circ\mathfrak{L}(x),\ \ \mu > 0, \qquad (2)$$

$$\text{where} \qquad \Psi_B(\cdot) := \Psi(\cdot) - \min_{v\in\mathcal{Z}}\left[\Psi(v) + \frac{1}{2}\|B(\cdot - v)\|^2\right].$$

Q1. Can we establish any iterative algorithm of guaranteed convergence to globally optimal solution of (2) under as much general overall-convexity conditions as possible ?

Although Ψ_B is nonsmooth and nonconvex, under mild conditions, we can express the set of all globally optimal solutions in terms of the fixed-point set of computable nonexpansive operator in a certain Hilbert space and therefore can solve (2).

J. Abe, M. Yamagishi, I. Yamada,
"Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition," Inverse Problems, (36pp), 2020.

Theorem 1 **Assume** $\dim(\mathcal{X}) < \infty$, $\dim(\mathcal{Z}) < \infty$,
$\Psi \in \Gamma_0(\mathcal{X})$ satisfies $\Psi \circ (-\mathrm{Id}) = \Psi$ & $A^*A - \mu \mathfrak{L}^* B^* B \mathfrak{L} \succeq O \, (\Rightarrow J_{\Psi_B \circ \mathfrak{L}} \in \Gamma_0(\mathcal{X}))$.

Define $T_{\mathrm{LiGME}} : \mathcal{X} \times \mathcal{Z} \times \mathcal{Z} \to \mathcal{X} \times \mathcal{Z} \times \mathcal{Z} : (x, v, w) \mapsto (\xi, \zeta, \eta)$ **by**

$$\xi := \left[\mathrm{Id} - \frac{1}{\sigma}(A^*A - \mu \mathfrak{L}^* B^* B \mathfrak{L})\right] x - \frac{\mu}{\sigma}\mathfrak{L}^* B^* B v - \frac{\mu}{\sigma}\mathfrak{L}^* w + \frac{1}{\sigma} A^* y$$

$$\zeta := \mathrm{Prox}_{\frac{\mu}{\tau}\Psi}\left[\frac{2\mu}{\tau} B^* B \mathfrak{L}\xi - \frac{\mu}{\tau} B^* B \mathfrak{L} x + \left(\mathrm{Id} - \frac{\mu}{\tau} B^* B\right) v\right]$$

$$\eta := \mathrm{Prox}_{\Psi^*}\left(2\mathfrak{L}\xi - \mathfrak{L}x + w\right),$$

where $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ **is chosen to satisfy**

$$\begin{cases} \sigma \mathrm{Id} - \frac{\kappa}{2} A^*A - \mu \mathfrak{L}^* \mathfrak{L} \succ O \\ \tau \geq \left(\frac{\kappa}{2} + \frac{2}{\kappa}\right) \mu \|B^* B\|_2. \end{cases}$$

$$\Psi^*(u) := \sup_{z \in \mathcal{Z}} \left(\langle z, u \rangle - \Psi(z)\right)$$

$$\mathrm{Prox}_{\Psi} + \mathrm{Prox}_{\Psi^*} = \mathrm{Id}$$

$$\operatorname{argmin} J_{\Psi_B \circ \mathfrak{L}} = \mathcal{Q}_{\mathcal{X}}\left(\mathrm{Fix}(T_{\mathrm{LiGME}})\right),$$

where $\mathcal{Q}_{\mathcal{X}} : \mathcal{X} \times \mathcal{Z} \times \mathcal{Z} \to \mathcal{X} : (x, v, w) \mapsto x$
and $\mathrm{Fix}(T_{\mathrm{LiGME}}) := \{(x, v, w) \mid T_{\mathrm{LiGME}}(x, v, w) = (x, v, w)\}$

$$\mathcal{P} := \begin{bmatrix} \sigma \mathrm{Id} & -\mu \mathfrak{L}^* B^* B & -\mu \mathfrak{L}^* \\ -\mu B^* B \mathfrak{L} & \tau \mathrm{Id} & O \\ -\mu \mathfrak{L} & O & \mu \mathrm{Id} \end{bmatrix} \succ O$$ **and**

$T_{\mathrm{LiGME}} : \mathcal{H} \, (:= \mathcal{X} \times \mathcal{Z} \times \mathcal{Z}) \to \mathcal{H}$ **is** $\frac{\kappa}{2\kappa - 1}$**-averaged nonexpansive in the Hilbert space** $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{P}}, \| \cdot \|_{\mathcal{P}})$**, i.e.,**

$(\forall \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{H})$

$$\| T_{\mathrm{LiGME}}(\mathbf{z}_1) - T_{\mathrm{LiGME}}(\mathbf{z}_2) \|_{\mathcal{P}}^2 \leq \| \mathbf{z}_1 - \mathbf{z}_2 \|_{\mathcal{P}}^2 - \frac{\kappa - 1}{\kappa} \| (\mathrm{Id} - T_{\mathrm{LiGME}})(\mathbf{z}_1) - (\mathrm{Id} - T_{\mathrm{LiGME}})(\mathbf{z}_2) \|_{\mathcal{P}}^2$$

For any initial point $(x_0, v_0, w_0) \in \mathcal{X} \times \mathcal{Z} \times \mathcal{Z}$,
the sequence $(x_n, v_n, w_n)_{n \in \mathbb{N}} \subset \mathcal{X} \times \mathcal{Z} \times \mathcal{Z}$ **generated by**

[Krasnoselskiĭ-Mann]

$$(x_{n+1}, v_{n+1}, w_{n+1}) := T_{\mathrm{LiGME}}(x_n, v_n, w_n)$$

converges to a point $(x^\star, v^\star, w^\star) \in \mathrm{Fix}\,(T_{\mathrm{LiGME}})$ **and**

$$\lim_{n \to \infty} x_n = x^\star \in \operatorname{argmin} \; J_{\Psi_B \circ \mathfrak{L}}$$

Contents of this talk

1.Why fixed pont expressions of closed convex sets ?

2. A fixed point strategy for
sparsity-aware estimation with nonconvex regularizations

3. A fixed point strategy for
Hierarchical Convex Optimization

Some variations of Proximal Splitting Algorithm

(e.g., Forward backward splitting / Primal-dual splitting / Douglas-Rachford splitting / ADMM ...)

can be seen as applications of

[Krasnosel'skiĭ('55)-Mann('53)]

$$u_{n+1} := (1-\alpha_n)\,u_n + \alpha_n T(u_n) \;\rightharpoonup \exists \widehat{u} \in \mathrm{Fix}(T)$$

to

Beautiful Expressions as

$$\mathcal{S}_p := \arg\min_{x\in\mathcal{X}} f(x) + g\circ A(x) = \Xi\,(\mathrm{Fix}(T))\,,$$

Proximal splitting Operators

where

$T : \mathcal{H} \to \mathcal{H}$: computable nonexpansive operators on $\mathcal{H}$

$\Xi : \mathcal{H} \to \mathcal{X}$: a bounded linear oprtaor

Most proximal splitting algorithms rely on K-M algorithm and can achieve convergence to only one anonymous solution, i.e.,

some $x^\star \in \mathcal{S}_p := \arg\min_{x \in \mathcal{X}} f(x) + g \circ A(x) = \Xi(\mathrm{Fix}(T))$

→ other solutions in $\mathcal{S}_p \setminus \{x^\star\}$ remain unavailable

and can utilize only a little information on $\mathcal{S}_p$!

This situation is wasting almost all valuable vectors in $\mathcal{S}_p$!

Q2. Can we choose a further best solution with new $\Psi \in \Gamma_0(\mathcal{X})$ by

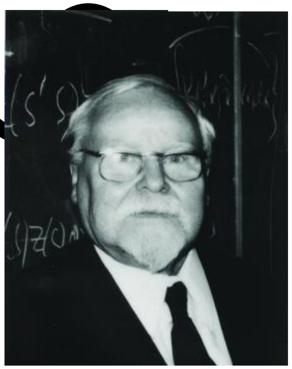
Hierarchical Convex Optimization ?

Minimize	$\Psi(x^\star)$	⇐ **2nd stage optimization**
Subject to	$x^\star \in \mathcal{S}_p$	⇐ **The set of all solutions of 1st stage optimization**

A central target in **Bilevel Optimization** (e.g., [Dempe, Zemkoho 2020]) !

A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," Soviet Math. Dokl., 4, 1963.

A Landmark Theorem of Tikhonov Regularization



Andrey Tikhonov (1906-1993) Moscow State University

Suppose that $\arg\min \varphi \cap \operatorname{dom}\Psi \neq \varnothing$

Ψ: **coercive and strictly convex.**

Let $x_\varepsilon \in \arg\min_{x \in \mathcal{X}} [\varphi(x) + \varepsilon\Psi(x)]$ **for every** $\varepsilon > 0$.

$$\left.\begin{aligned} x_\varepsilon &\rightharpoonup x^{\star\star} \in \arg \min_{x^\star \in \arg\min \varphi} \Psi(x^\star) \\ \Psi(x_\varepsilon) &\to \min_{x^\star \in \arg\min \varphi} \Psi(x^\star) \end{aligned}\right\} \quad (\text{as } \varepsilon \downarrow 0)$$

suggests computational difficulty in the Hierarchical Convex Optimization !

Fortunately,

we can plug many Proximal Splitting Operators into Hybrid Steepest Descent Method for Hierarchical Convex Optimization !

(P1)

To minimize $\Psi(x^{\star})$ subject to $x^{\star} \in \mathcal{S}_p$,

we have found many practical ways:

[Yamada-Ogura-Shirakawa '02],[Yamada-Yukawa-Yamagishi '11], [Ono-Yamada '14], [Yamagishi-Yamada '17], [Yamada-Yamagishi'19]

by exploiting specially nice expressions

$$\mathcal{S}_p := \arg\min_{x \in \mathcal{X}} f(x) + g \circ A(x) = \Xi\left(\mathrm{Fix}(T)\right),$$

where $\Xi : \mathcal{H} \to \mathcal{X}$ is a bounded linear operator,

and by translating (P1) into

(P2)

Minimize $\Theta := \Psi \circ \Xi$ over $\mathrm{Fix}(T) \subset \mathcal{H}$

A Key for Hierarchical Convex Optimization

Hybrid Steepest Descent Method

[Yamada et al '96, Deutsch-Yamada'98, Yamada'01, Ogura-Yamada'03, ...]

$$u_{n+1} := T(u_n) - \lambda_{n+1} \nabla\Theta\left(T(u_n)\right)$$

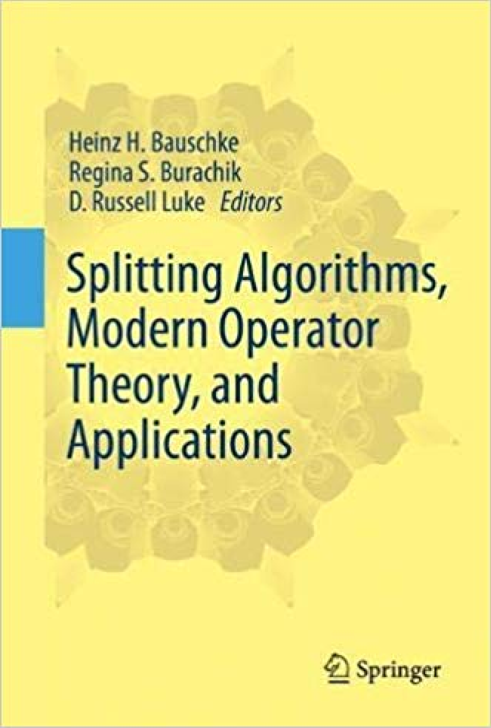
can minimize $\Theta(:= \Psi \circ \Xi)$ over

$$\mathrm{Fix}(T) := \{u \in \mathcal{H} \mid T(u) = u\}$$

where

$$\begin{cases} \Theta : \mathcal{H} \to \mathbb{R}, & \text{Smooth Convex Function} \\ \nabla\Theta : \mathcal{H} \to \mathcal{H}, & \text{Lipschitz Continuous} \\ T : \mathcal{H} \to \mathcal{H}, & \text{Nonexpansive operator} \\ (\lambda_n)_{n=1}^{\infty} \subset [0, \infty) : \text{slowly decreasing, s.t.,} & \begin{cases} \lim_{n\to\infty} \lambda_n = 0, \\ \sum_{n=1}^{\infty} \lambda_n = \infty \end{cases} \end{cases}$$

1. This is extension of [Halpern'67/ Reich'74 / Lions'77/ Wittmann'92/...]
2. This scheme achieves a very best vector among all fixed points !



For details of *Hierarchical Convex Optimization by HSDM+Proximal Splitting Operators*

I. Yamada, M. Yamagishi, Hierarchical Convex Optimization by the Hybrid Steepest Descent Method with Proximal Splitting Operators - Enhancements of SVM and Lasso,

In : H. H. Bauschke, R. Burachik and D. R. Luke eds.,
Splitting Algorithms, Modern Operator Theory, and Applications,
pp.413-489, Springer, 2019.

For an application to multiclass SVM, see

Y. Nakayama, M. Yamagishi, I. Yamada,
"A hierarchical convex optimization for multiclass SVM achieving maximum pairwise margins with least empirical hinge-loss,"
arXiv2004.08180, 2020.

Conclusion

Related info is found in the following papers.

1. Linearly involved Generalized Moreau Enhanced (LiGME) models for sparsity-rank-aware signal processing

J. Abe, M. Yamagishi, I. Yamada, "Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition," Inverse Problems, (36pp), 2020.

Y. Zhang, I. Yamada, "A unified class of DC-type convexity preserving regularizers for improved sparse regularization," EUSIPCO 2022.

W.Yata, M.Yamagishi, I.Yamada, "A constrained LiGME Model and Its Proximal Splitting Algorithm under overall convexity condition," J. Applied and Numerical Optimization," 2022.

2. Hierarchical convex optimization by hybrid steepest descent method with proximal splitting operators

I. Yamada, "The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings," In: D. Butnariu et al. eds., Inherently Parallel Alg. in Feasibility and Optimization and Their Applications, pp. 473–504. Elsevier, 2001.

I. Yamada, M. Yamagishi, "Hierarchical Convex Optimization by the Hybrid Steepest Descent Method with Proximal Splitting Operators - Enhancements of SVM & Lasso," In : H.H.Bauschke, et al eds., Splitting Algorithms, Modern Operator Theory, and Applications, pp.413-489, Springer, 2019.

3. A comprehensive tutorial on fixed point strategies in data science

P. L.Combettes, J.C.Pesquet, Fixed point strategies in data science, IEEE Trans Signal Process, vol.69, pp.3878-3905, 2021.