Hierarchical Convex Optimization with Proximal Splitting Operators

Isao Yamada
Tokyo Institute of Technology

Based on joint work with Masao Yamagishi

Dedicated to the memory of Jonathan M. Borwein

Splitting Algorithms, Modern Operator Theory, & Applications
September 19, 2017, Casa Matemática Oaxaca

Convex Optimization - A typical form :

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

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

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

has been playing a central role in Inverse Problems because

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

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

can be handled as an instance of (P) by

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

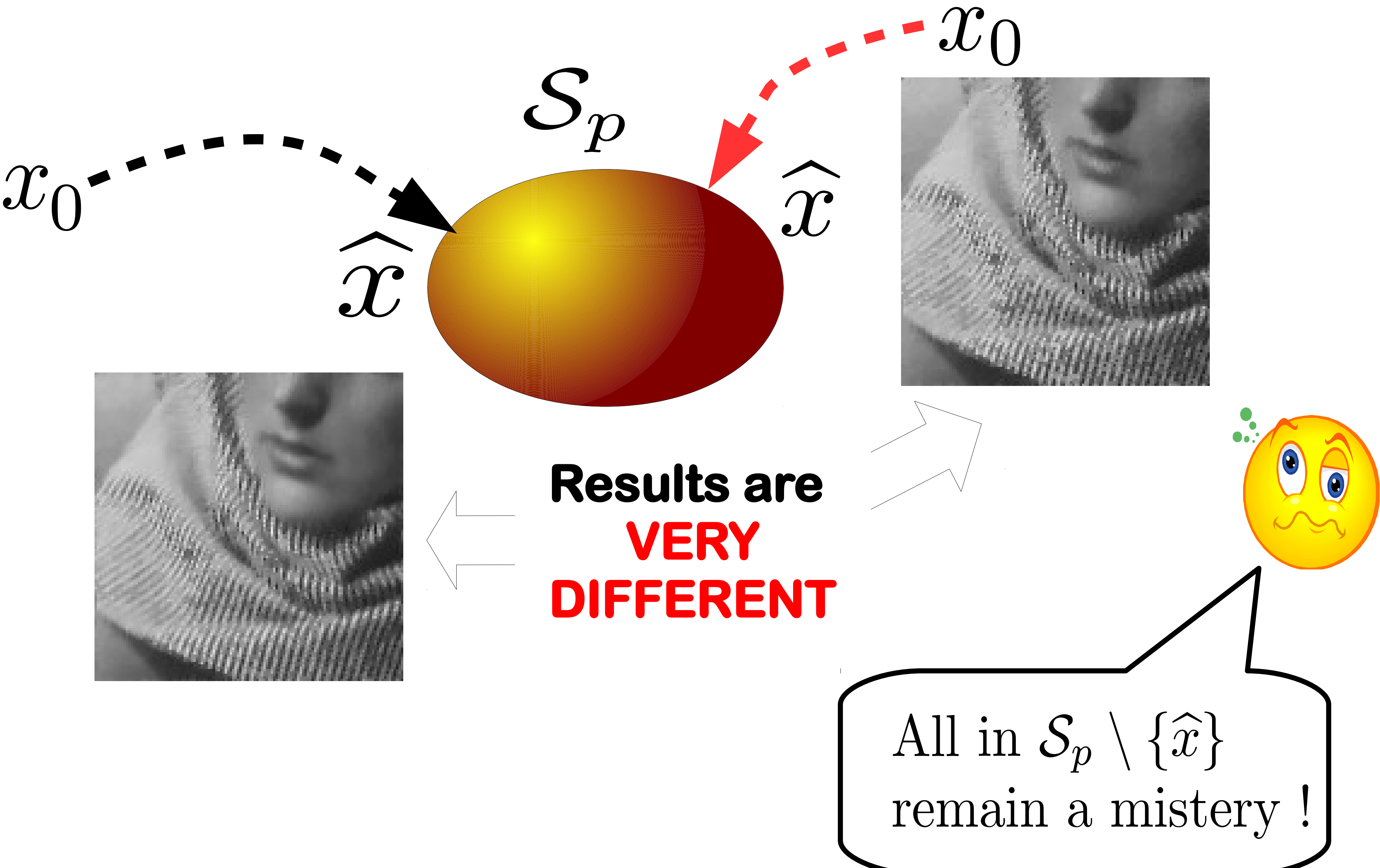
But

Almost all existing algorithms achieve convergence to only one unspecial solution :

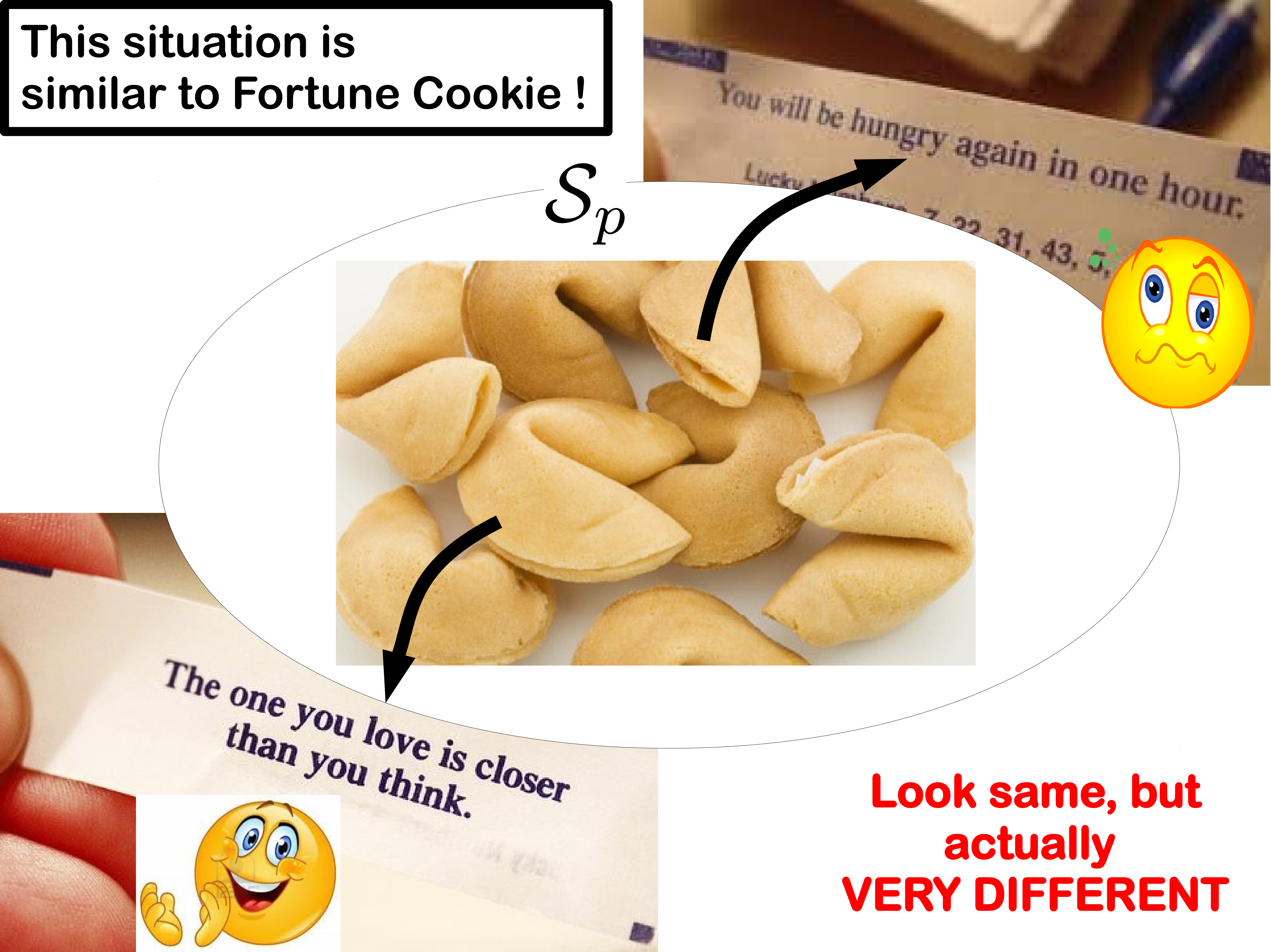
$$x^\star \in \mathcal{S}_p := \arg\min_{x \in \mathcal{X}} f(x) + g(Ax) \neq \emptyset.$$

Other solutions in $\mathcal{S}_p \setminus \{x^\star\}$ remain mistery !

Imagine, e.g., convex feasibility problems !


x_0
$\mathcal{S}_p$
x_0
$\widehat{x}$
$\widehat{x}$
Results are
VERY
DIFFERENT
All in $\mathcal{S}_p \setminus \{\widehat{x}\}$
remain a mistery !

This situation is similar to Fortune Cookie !



Look same, but actually VERY DIFFERENT

Challenges for strategic convergence are found, e.g,

Better limit

Superiorization:

[Censor-Davidi-Herman '10], [Herman-Garduno-Davidi-Censor '12]

An idea to incorporate a faborable attribute into a given iterative algorithm, without changing the inherent desired properties of the algorithm.

We are trying to find Best limit :

Hierarchical convex optimization :

[Yamada-Ogura-Shirakawa '02], [Yamada-Yukawa-Yamagishi '11], [Ono-Yamada '15], [Yamagishi-Yamada '17]

Best limit in what sense ?

Hierarchical Convex Optimization

Suppose $\mathcal{S}_p := \arg\min_{x \in \mathcal{X}} f(x) + g(Ax) \neq \varnothing$ &

$\Psi \in \Gamma_0(\mathcal{X})$ **is desired to be minimized additionally, i.e.,**

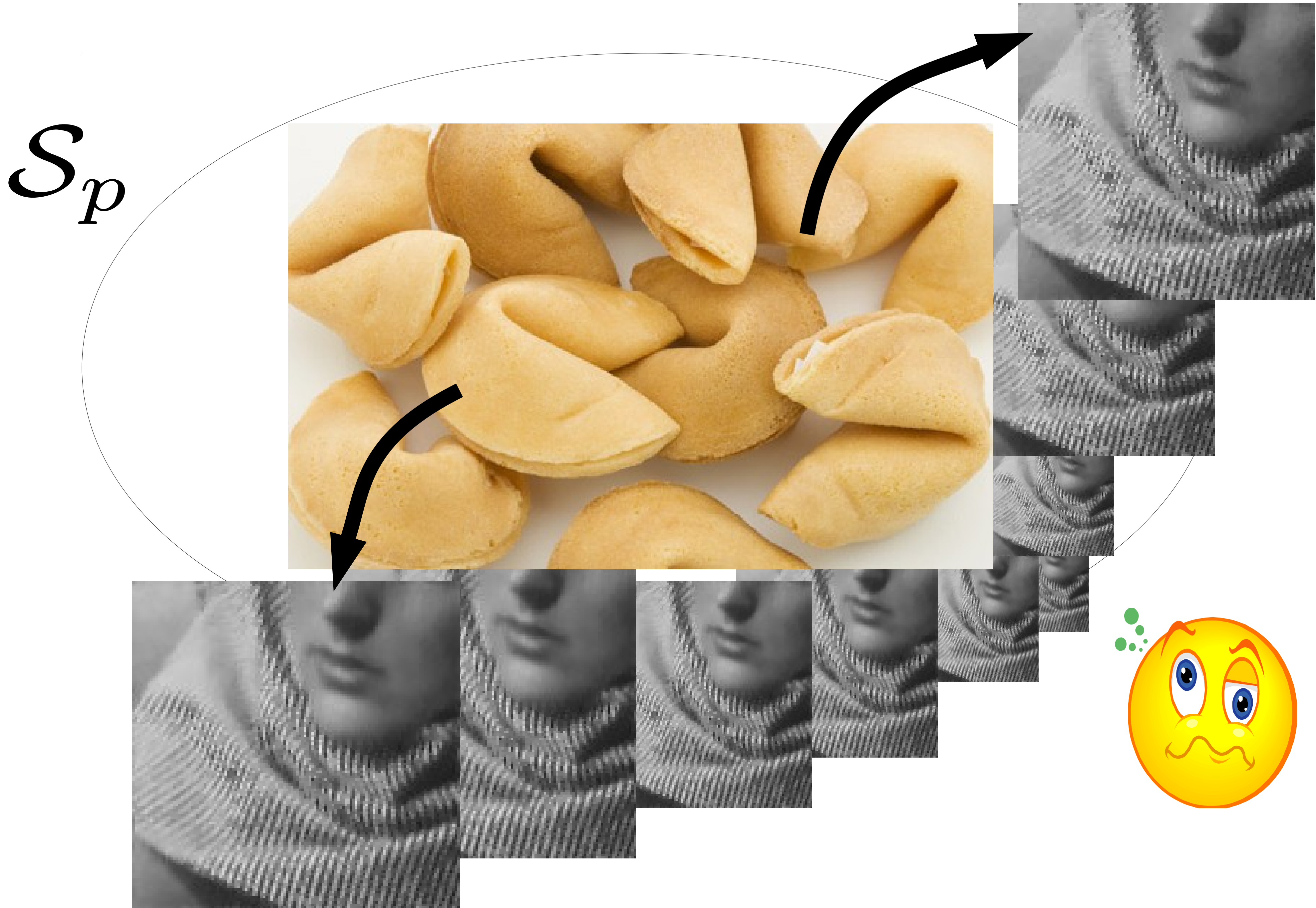
Minimize	$\Psi(x^\star)$	⇦ 2nd stage optimization
Subject to	$x^\star \in \mathcal{S}_p$	⇦ The set of all solutions of 1st stage optimization

For this challenging mission impossible, we need at least

1. Exploiting **Full information** on $\mathcal{S}_p$ (usually infinite set in $\mathcal{X}$).
2. Mathematically sound **algorithmic ideas** to minimize Ψ over $\mathcal{S}_p$.

Hierarchical convex optimization casts a question:

Can we choose a best one without crunching all cookies ?



PART I

Preliminaries :

How can we capture full information on the solution set of Convex Optimization Problem ?

PART II

Hierarchical Convex Optimization :

How can we choose a Best Fortune Cookie

without crunching all cookies ?

PART III

Application to

State-of-the-art Statistiacal Estimation Technique

A Hierarchical Enhancement of Lasso

Convex Optimization Problem

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

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

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

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	$(\forall x, y \in \mathrm{dom}\varphi, \forall \lambda \in (0,1))$ $\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)$

The solution sets of convex optimization problems can often be expressed as

$$\arg\min_{x \in \mathcal{X}} \varphi(x) = \Xi\left(\mathrm{Fix}(T)\right)$$

where

$T : \mathcal{H} \to \mathcal{H}$: a **nonexpansive operator** defined on a certain Hilbert space $\mathcal{H}$

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

$\Xi : \mathcal{H} \to 2^{\mathcal{X}}$: a certain set-valued operator

Computable Nonexpansive Operators for Convex Optimization

. . . Primal-Dual splitting Operator . . .

ADMM Operator (Dual Variant of Douglas-Rachford splitting operator)

Great !

Augmented Lagrangian Operator

Forward-Backward splitting operator

Proximity operator

$$\mathrm{prox}_f := (\mathrm{Id} + \partial f)^{-1} \quad (f \in \Gamma_0(\mathcal{H}))$$

$$: \mathbf{x} \mapsto \arg\min_{\mathbf{y} \in \mathcal{H}} \left[f(\mathbf{y}) + \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 \right]$$

Metric projection

Nonexpansive Operator

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

Proximity Operator (J.J.Moreau '62)

$f \in \Gamma_0(\mathcal{X})$

$$\mathrm{prox}_f : \mathcal{X} \to \mathcal{X} : x \mapsto \arg\min_{y \in \mathcal{X}} \left\{ f(y) + \frac{1}{2}\|x - y\|^2 \right\}$$

is $1/2$- averaged nonexpansive operator, i.e.,

$\mathrm{rprox}_f := 2\mathrm{prox}_f - \mathrm{Id}$ is nonexpansive.

$$\begin{aligned}
& z \in \arg\min_{x \in \mathcal{X}} f(x) \\
\Leftrightarrow \quad & 0 \in \partial f(z) := \{p \in \mathcal{X} \mid f(z) + \langle p, x - z \rangle \le f(x)\ (\forall x \in \mathcal{X})\} \in 2^{\mathcal{X}} \\
\Leftrightarrow \quad & z \in z + \partial f(z) = (\mathrm{Id} + \partial f)(z) \in 2^{\mathcal{X}} \\
\Leftrightarrow \quad & z = (\mathrm{Id} + \partial f)^{-1}(z) = \mathrm{prox}_f(z) \\
\Leftrightarrow \quad & z \in \mathrm{Fix}\left(\mathrm{prox}_f\right)
\end{aligned}$$

Subdifferential of f at z

Proximity operator of f

Resolvent of ∂f

Proximity Operator of Conjugate function

$$\forall f \in \Gamma_0(\mathcal{X}),\ f^*: \mathcal{X} \ni y \mapsto \sup_{x \in \mathcal{X}} \left(\langle y, x \rangle - f(x) \right) \in (-\infty, \infty]$$

is called **Fenchel-Rockafellar Conjugate** of f

and satisfies $f^* \in \Gamma_0(\mathcal{X})$ &

Inverse Resolvent Identity

$$\mathrm{Id} = \mathrm{prox}_f + \mathrm{prox}_{f^*}$$

If $f \in \Gamma_0(\mathcal{X})$ is **prox-friendly** (i.e., prox_f is easily computable), $f^* \in \Gamma_0(\mathcal{X})$ is also **prox-friendly**.

Most splitting algorithms more or less rely on ...

Fact (Krasnosel'skii-Mann, e.g.[Mann'53,Dotson'70,Groetsch'72])

Suppose $T : \mathcal{H} \to \mathcal{H}$ is $\begin{cases} \text{Nonexpansive} \\ \mathrm{Fix}(T) \neq \emptyset \end{cases}$

Then for any $\begin{cases} \forall x_0 \in \mathcal{H} \\ (\alpha_n)_{n=0}^{\infty} \subset (0,1) \text{ s.t. } \sum_{n=0}^{\infty} \alpha_n (1-\alpha_n) = \infty \end{cases}$

$$x_{n+1} := (1-\alpha_n) x_n + \alpha_n T(x_n) \rightharpoonup \exists \hat{x} \in \mathrm{Fix}(T)$$

In fact, after careful observations, we can interpret

Proximal Splitting Algorithms
(Forward backward splitting/Primal-dual splitting/ Douglus -Rachford splitting / ADMM etc)

as applications of **K-M Alg**

to

$$\arg\min_{x \in \mathcal{X}} f(x) + g(Ax) = \Xi(\mathrm{Fix}(T))$$

where

$T : \mathcal{H} \to \mathcal{H}$: a computable nonexpansive operator defined on a certain Hilbert space $\mathcal{H}$

$\Xi : \mathcal{H} \to 2^{\mathcal{X}}$: a certain set-valued operator

Example (ADMM e.g. [Gabay '83])

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

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

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

$$\begin{cases} x_{k+1} \in \underset{x \in \mathcal{X}}{\arg\min} \left(f(x) + \frac{1}{2} \|Ax - y_k - \nu_k\|_{\mathcal{K}}^2 \right) \\ y_{k+1} \in \underset{y \in \mathcal{K}}{\arg\min} \left(g(y) + \frac{1}{2} \|Ax_{k+1} - y - \nu_k\|_{\mathcal{K}}^2 \right) \\ \nu_{k+1} = \nu_k - Ax_{k+1} + y_{k+1} \end{cases}$$

A Fixed Point Theoretic View of ADMM [Eckstein-Bertsekas'92]

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

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

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

$$\text{(D)} \quad \underset{u\in\mathcal{K}}{\text{minimize}}\ f^*(A^*u) + g^*(-u)$$

$$\theta_1 := f^* \circ A^*,\ \theta_2 := g^* \circ (-\mathrm{Id})$$

$$\mathcal{S}_d := \arg\min_{\nu\in\mathcal{K}} f^*(A^*\nu) + g^*(-\nu) = \mathrm{prox}_{\theta_2}\left(\mathrm{Fix}\left(\mathrm{rprox}_{\theta_1}\mathrm{rprox}_{\theta_2}\right)\right)$$

$$\mathcal{S}_p := \arg\min_{x\in\mathcal{X}} f(x) + g(Ax) = \partial f^*(A^*\nu^\star) \cap A^{-1}\left(\partial g^*(-\nu^\star)\right) \quad (\forall \nu^\star \in \mathcal{S}_d)$$

$$\text{ADMM} = \text{K-M alg for a point in } \mathrm{Fix}(\mathrm{rprox}_{\theta_1}\mathrm{rprox}_{\theta_2})$$

PART I

Preliminaries :

How can we capture full information on the solution set of Convex Optimization Problem ?

PART II

Hierarchical Convex Optimization :

How can we choose a Best Fortune Cookie without crunching all cookies ?

PART III

Application to State-of-the-art Statistiacal Estimation Technique

A Hierarchical Enhancement of Lasso

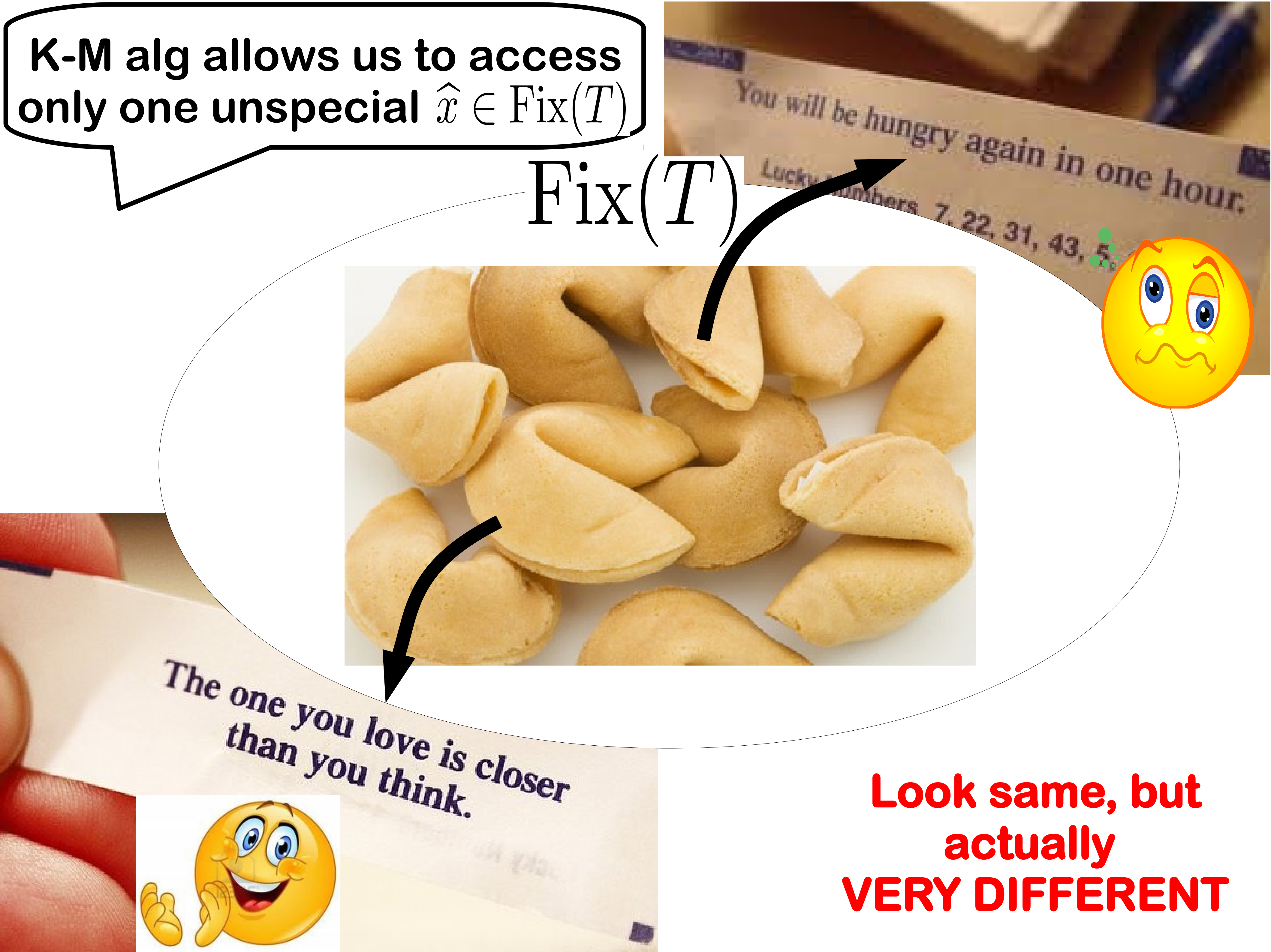
We have seen the solution sets of convex optimization problems can often be expressed as

$$\arg\min_{x \in \mathcal{X}} f(x) + g(Ax) = \Xi\left(\mathrm{Fix}(T)\right)$$

where

$T : \mathcal{H} \to \mathcal{H}$: a **computable nonexpansive operator** defined on a certain Hilbert space $\mathcal{H}$

$\Xi : \mathcal{H} \to 2^{\mathcal{X}}$: a certain set-valued operator


K-M alg allows us to access only one unspecial $\widehat{x} \in \mathrm{Fix}(T)$
You will be hungry again in one hour.
$\mathrm{Fix}(T)$
The one you love is closer than you think.
Look same, but actually VERY DIFFERENT

Can we choose best one without crunching all cookies ?

This mission
can not be accomplishd by K-M alg
but could be accomplished by
Hybrid steepest descent method !

K-M algorithm

$$x_{n+1} := (1-\alpha_n)x_n + \alpha_n T(x_n) \rightharpoonup \exists \hat{x} \in \mathrm{Fix}(T)$$

Hybrid Steepest Descent Method

$$x_{n+1} := T(x_n) - \lambda_{n+1} \nabla \Psi\left(T(x_n)\right) \to \text{Best Point} \in \mathrm{Fix}(T)$$

A Key for Hierarchical Convex Optimization

Hybrid Steepest Descent Method

[Yamada et al '96, Deutsch-Yamada'98, Yamada'01, Yamada-Ogura'04 etc]

$$x_{n+1} := T(x_n) - \lambda_{n+1} \nabla \Psi \left(T(x_n) \right)$$

can minimize Ψ over

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

where

$\Psi : \mathcal{H} \to \mathbb{R}$, Smooth Convex Function

$\nabla \Psi : \mathcal{H} \to \mathcal{H}$, Lipschitz Continuous

$T : \mathcal{H} \to \mathcal{H}$, Nonexpansive operator

$(\lambda_n)_{n=1}^{\infty} \subset [0, \infty)$: Slowly decreasing

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

Theorem (Convergence of HSDM, see, e.g. [Yamada'01])

$T : \mathcal{H} \to \mathcal{H}$ Nonexpansive with $\mathrm{Fix}(T) \neq \varnothing$

$\Psi : \mathcal{H} \to \mathbb{R}$ Gâteaux differentiable s.t.

$$(\exists \kappa, \eta > 0, \forall x, y \in T(\mathcal{H})) \quad \begin{aligned} &\|\nabla\Psi(x) - \nabla\Psi(y)\| \leq \kappa \|x - y\| \\ &\langle \nabla\Psi(x) - \nabla\Psi(y), x - y \rangle \geq \eta \|x - y\|^2 \end{aligned}$$

$$(\lambda_n)_{n\geq 1} \subset [0, \infty) \text{ satisfies } \begin{cases} \text{(i)} & \lim_{n\to\infty} \lambda_n = 0 \\ \text{(ii)} & \sum_{n\geq 1} \lambda_n = \infty \\ \text{(iii)} & \sum_{n\geq 1} |\lambda_n - \lambda_{n+1}| < \infty \end{cases}$$

$$(\forall x_0 \in \mathcal{H}) \quad x_{n+1} := T(x_n) - \lambda_{n+1} \nabla\Psi(T(x_n))$$

satisfies $\displaystyle \lim_{n\to\infty} \|x_n - x^{\star\star}\| = 0$

where $\displaystyle x^{\star\star} \in \Omega := \arg \min_{x \in \mathrm{Fix}(T)} \Psi(x)$ (Note: $|\Omega| = 1$)

Theorem (nonstrictly convex, $\dim(\mathcal{H}) < \infty$ [Ogura-Yamada'03])

Suppose

$T : \mathcal{H} \to \mathcal{H}$ Nonexpansive with bounded $\mathrm{Fix}(T) \neq \emptyset$

$\Psi : \mathcal{H} \to \mathbb{R}$ Smooth Convex function, s.t.

$$(\exists \kappa > 0, \forall x, y \in T(\mathcal{H})) \quad \|\nabla\Psi(x) - \nabla\Psi(y)\| \leq \kappa\|x - y\|$$

$$(\lambda_n)_{n\geq 0} \in \ell^2_+ \setminus \ell^1_+.$$

Then

$$(\forall x_0 \in \mathcal{H}) \quad x_{n+1} := T(x_n) - \lambda_{n+1}\nabla\Psi(T(x_n))$$

satisfies $$\lim_{n\to\infty} \mathrm{d}(x_n, \Omega) = 0$$

where $$\Omega := \arg \min_{x\in\mathrm{Fix}(T)} \Psi(x) \neq \emptyset.$$

How can we combine Nonexpansive Operators with Hybrid Steepest Descent Method for Hierarchical Convex Optimization ?

We have found many ways ! See for example

[Yamada-Ogura-Shirakawa '02],[Yamada-Yukawa-Yamagishi '11],
[Ono-Yamada '15], [Yamagishi-Yamada '17]

Next we demonstrate a simple strategy in an application to statistical estimation problem !

PART I

Preliminaries :

How can we capture full information on
the solution set of Convex Optimization Problem ?

PART II

Hierarchical Convex Optimization :

How can we choose a Best Fortune Cookie
without crunching all cookies ?

PART III

Application to
State-of-the-art Statistiacal Estimation Technique

A Hierarchical Enhancement of Lasso

Consider the estimation of $\mathbf{b}^{\mathrm{tru}} \in \mathbb{R}^p$ in the standard linear model:

$$\mathbf{z} = \mathbf{X}\mathbf{b}^{\mathrm{tru}} + \sigma\mathbf{e}$$

where

Response vector

$$\mathbf{z} = (z_1, z_2, \ldots, z_n)^t \in \mathbb{R}^n$$

Design matrix

$\mathbf{X} \in \mathbb{R}^{n \times p}$ ($p > n$ for High-dimensional case)

: assumed to have no zero column vector !

Noise vector

$$\mathbf{e} = (\varepsilon_1, \ldots, \varepsilon_n)^t \in \mathbb{R}^n$$

ε_i : realization of normalized random variable with mean 0 and variance 1

Standard Deviation of Entire noise

$$\sigma > 0$$

Lasso [Robert Tibshirani '96]

$$\widehat{\mathbf{b}}_{\text{Lasso}}(\lambda) \in \arg\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \frac{1}{2n}\|\mathbf{z} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1 \right\}.$$

A Prediction Bound for Lasso

[Koltchinskii, Lounici, and Tsybakov'11], [Rigollet and Tsybakov'11]

If $\lambda \geq \dfrac{2\|\mathbf{X}^t(\mathbf{z} - \mathbf{X}\mathbf{b}^{\text{tru}})\|_\infty}{n}$,

it holds $\dfrac{\|\mathbf{X}\widehat{\mathbf{b}}_{\text{Lasso}}(\lambda) - \mathbf{X}\mathbf{b}^{\text{tru}}\|_2^2}{n} \leq 2\lambda\|\mathbf{b}^{\text{tru}}\|_1.$

A powerful enhancement of Lasso

TREX [Lederer, Müller '15]: Nonconvex Optimization

$$\widehat{\mathbf{b}}_{\text{TREX}} \in \arg\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \frac{\|\mathbf{X}\mathbf{b} - \mathbf{z}\|_2^2}{\|\mathbf{X}^t(\mathbf{X}\mathbf{b} - \mathbf{z})\|_\infty} + \alpha\|\mathbf{b}\|_1 \right\}, \text{default choice } \alpha = \frac{1}{2}$$

TREX: an enhancement of Lasso [Lederer, Müller '15]

Nonconvex

$$\widehat{\mathbf{b}}_{\mathrm{TREX}} \in \arg\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \frac{\|\mathbf{X}\mathbf{b}-\mathbf{z}\|_2^2}{\alpha\|\mathbf{X}^t(\mathbf{X}\mathbf{b}-\mathbf{z})\|_\infty} + \|\mathbf{b}\|_1 \right\}, \quad \text{default choice } \alpha = \frac{1}{2}$$

where $\|\mathbf{X}^t(\mathbf{X}\mathbf{b}-\mathbf{z})\|_\infty = \max_{1\le j\le p} |\mathbf{X}^t_{:j}(\mathbf{X}\mathbf{b}-\mathbf{z})|$

Great News 1

[Bien, Gaynanova, Lederer, and Müller '16]

Find best one among solutions of $2p$ convex subproblems:

$$\widehat{\mathbf{b}}^{(j)}_{\mathrm{TREX}} \in \operatorname*{argmin}_{\substack{\mathbf{b}\in\mathbb{R}^p \\ \mathbf{x}_j^t(\mathbf{X}\mathbf{b}-\mathbf{z})>0}} \left\{ \frac{\|\mathbf{X}\mathbf{b}-\mathbf{z}\|_2^2}{\alpha\mathbf{x}_j^t(\mathbf{X}\mathbf{b}-\mathbf{z})} + \|\mathbf{b}\|_1 \right\}$$

where $\mathbf{x}_j = \pm\mathbf{X}_{:j} \quad (j = 1, 2, \ldots, p)$

j th convex subproblem of TREX [Bien, Gaynanova, Lederer, and Müller '16]

$$\widehat{\mathbf{b}}^{(j)}_{\mathrm{TREX}} \in \underset{\substack{\mathbf{b} \in \mathbb{R}^p \\ \mathbf{x}_j^t(\mathbf{X}\mathbf{b}-\mathbf{z}) > 0}}{\operatorname{argmin}} \left\{ \frac{\|\mathbf{X}\mathbf{b}-\mathbf{z}\|_2^2}{\alpha \mathbf{x}_j^t(\mathbf{X}\mathbf{b}-\mathbf{z})} + \|\mathbf{b}\|_1 \right\}$$

A Reformulation for Proximal Splitting [Combettes, Müller '17]

$$\widehat{\mathbf{b}}^{(j)}_{\mathrm{pTREX}} \in \mathcal{S}_j := \underset{\mathbf{b}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ g_j(\mathbf{M}_j \mathbf{b}) + \|\mathbf{b}\|_1 \right\},$$

Great News 2

$$g_j : \mathbb{R} \times \mathbb{R}^n \to (-\infty, \infty] : (\eta, \mathbf{y}) \mapsto \begin{cases} \frac{\|\mathbf{y}-\mathbf{z}\|_2^2}{\alpha(\eta - \mathbf{x}_j^t \mathbf{z})}, & \text{if } \eta > \mathbf{x}_j^t \mathbf{z}; \\ 0, & \text{if } \mathbf{y} = \mathbf{z} \text{ and } \eta = \mathbf{x}_j^t \mathbf{z}; \\ +\infty, & \text{otherwise.} \end{cases}$$

$$\in \Gamma_0(\mathbb{R} \times \mathbb{R}^n) \; : \; \text{proper lower-semicontinuous convex,}$$

$$\operatorname{prox}_{g_j} := (\mathrm{Id} + \partial g_j)^{-1} \; : \; \text{has a closed form expression,}$$

$$\mathbf{M}_j : \mathbb{R}^p \to \mathbb{R} \times \mathbb{R}^n : \mathbf{b} \mapsto (\mathbf{x}_j^t \mathbf{X}\mathbf{b}, \mathbf{X}\mathbf{b}) \; : \; \text{Bounded Linear}$$

Product Space Reform. TREX subproblem [Combettes, Müller '17]

$$\underset{\mathbf{b}\in\mathbb{R}^p}{\text{minimize}} \quad g_j(\mathbf{M}_j\mathbf{b}) + \|\mathbf{b}\|_1 \qquad \text{Convex Optimization over } \mathbb{R}^p$$

$$\underset{\mathbf{x}=(\mathbf{b},\mathbf{c})\in\mathbb{R}^p\times\mathbb{R}^{n+1}}{\text{minimize}} \quad F_j(\mathbf{x}) + G_j(\mathbf{x}) \qquad \text{Convex Optimization over } \mathbb{R}^{p+n+1}$$

$$F_j : (\mathbf{b},\mathbf{c}) \mapsto \|\mathbf{b}\|_1 + g_j(\mathbf{c}), \qquad G_j(\mathbf{b},\mathbf{c}) = \begin{cases} 0 & \text{if } \mathbf{M}_j\mathbf{b} = \mathbf{c}, \\ \infty & \text{otherwise} \end{cases}$$

$$\text{prox}_{F_j}(\mathbf{b},\mathbf{c}) = \left(\text{prox}_{\|\cdot\|_1}(\mathbf{b}), \text{prox}_{g_j}(\mathbf{c})\right) \qquad \text{prox}_{G_j}(\mathbf{b},\mathbf{c}) = \text{projection onto}$$

soft thresholder

$$\mathcal{M}_j := \{(\mathbf{b},\mathbf{c}) \in \mathbb{R}^p \times \mathbb{R}^{n+1} \mid \mathbf{M}_j\mathbf{b} = \mathbf{c}\}$$

A Fixed Point Characterization with Douglas-Rachford Operator

$$\mathbf{x}^\star \text{ minimizes } F_j + G_j \quad \Leftrightarrow \quad \begin{cases} \mathbf{x}^\star = \text{prox}_{\gamma G_j}(\mathbf{y}) = P_{\mathcal{M}_j}(\mathbf{y}) \\ \mathbf{y} \in \text{Fix}(T_j), \end{cases}$$

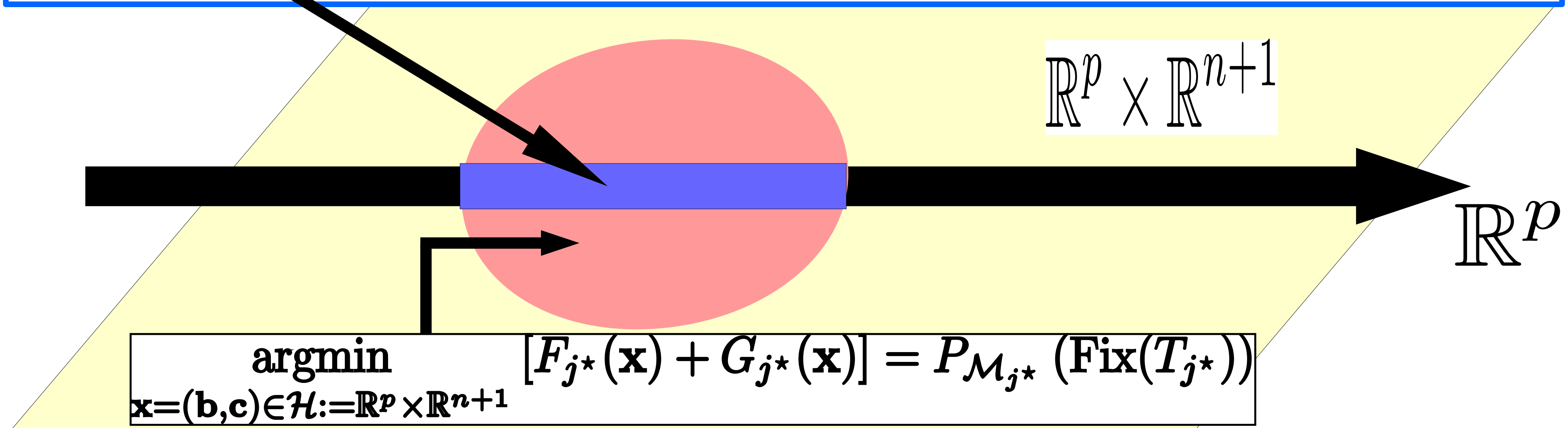
where $T_j := \left(2\text{prox}_{\gamma F_j} - \text{Id}\right) \circ \left(2P_{\mathcal{M}_j} - \text{Id}\right) : \mathbb{R}^{p+n+1} \to \mathbb{R}^{p+n+1}$: Nonexpansive

The solution set $\mathcal{S}$ of TREX can be expressed completely in terms of computable operators !

Solution set of **TREX**

$$\mathcal{S} := \underset{\mathbf{b}\in\mathbb{R}^p}{\operatorname{argmin}} \{g_{j^\star}(\mathbf{M}_{j^\star}\mathbf{b}) + \|\mathbf{b}\|_1\} = \mathcal{Q}_p \circ P_{\mathcal{M}_{j^\star}}(\operatorname{Fix}(T_{j^\star}))$$

where $\mathcal{Q}_p : \mathbb{R}^p \times \mathbb{R}^{n+1} \to \mathbb{R}^p : (\mathbf{b}, \mathbf{c}) \mapsto \mathbf{b}$



We have shown

The nonexpansive operator $T_{j^\star} := \left(2\operatorname{prox}_{\gamma F_{j^\star}} - \operatorname{Id}\right) \circ \left(2P_{\mathcal{M}_{j^\star}} - \operatorname{Id}\right)$ has **BOUNDED** $\operatorname{Fix}(T_{j^\star}) = \{\mathbf{y} \in \mathbb{R}^{p+n+1} \mid T_{j^\star}(\mathbf{y}) = \mathbf{y}\}$!

Consider Hierarchical Enhancement of TREX for

$$\mathbf{z} = \mathbf{X}\mathbf{b}^{\mathrm{tru}} + \sigma \mathbf{e}$$

TREX: Reformulation of convex subproblems [Combettes, Muller '17]

Find vectors in $\mathcal{S}_j := \underset{\mathbf{b} \in \mathbb{R}^p}{\operatorname{argmin}} \{g_j(\mathbf{M}_j \mathbf{b}) + \|\mathbf{b}\|_1\} \quad (j = 1, 2, , 2p),$

where $\quad g_j : \mathbb{R} \times \mathbb{R}^n \to (-\infty, \infty] : (\eta, \mathbf{y}) \mapsto \begin{cases} \frac{\|\mathbf{y}-\mathbf{z}\|_2^2}{\frac{1}{2}(\eta - \mathbf{x}_j^t \mathbf{z})}, & \text{if } \eta > \mathbf{x}_j^t \mathbf{z}; \\ 0, & \text{if } \mathbf{y} = \mathbf{z} \text{ and } \eta = \mathbf{x}_j^t \mathbf{z}; \\ +\infty, & \text{otherwise,} \end{cases}$

Further enhancement of TREX with prior knowledge ?

Example

Suppose that we know the target $\mathbf{b}^{\mathrm{tru}}$ is **fairly flat** !

$$\|D\mathbf{b}^{\mathrm{tru}}\|^2 \text{ is expected to be small,}$$

where $D : \mathbb{R}^p \to \mathbb{R}^{p-1}$ returns differences of all neighboring pairs.

A Hierarchical Convex Optimization for Enhancement of TREX

For $j = 1, 2, \ldots, 2p$,

Problem (Hierarchical enhancement of Lasso for promoting Flatness)

$$\left.\begin{array}{ll} \text{Minimize} & \left\| D\mathbf{b}^\star \right\|^2 \\ \text{subject to} & \mathbf{b}^\star \in \mathcal{S}_j := \underset{\mathbf{b}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ g_j(\mathbf{M}_j \mathbf{b}) + \|\mathbf{b}\|_1 \right\} \\ & \qquad\;\; = \mathcal{Q}_p \circ P_{\mathcal{M}_j}\left(\operatorname{Fix}(T_j)\right), \end{array}\right\} \quad (\star)$$

where

$$D := \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

(Equivalent Problem for **Hybrid Steepest Descent Method**)

$$\begin{array}{ll} \text{Minimize} & \Psi(\mathbf{x}^\star) := \left\| D \circ \mathcal{Q}_p \circ P_{\mathcal{M}_j}(\mathbf{x}^\star) \right\|^2 \\ \text{subject to} & \mathbf{x}^\star \in \operatorname{Fix}(T_j) \end{array} \quad (\star\star)$$

where $T_j := \left(2\operatorname{prox}_{\gamma F_j} - \mathrm{Id}\right) \circ \left(2P_{\mathcal{M}_j} - \mathrm{Id}\right) : \mathbb{R}^{p+n+1} \to \mathbb{R}^{p+n+1}$: Nonexpansive

Numerical Test (Underdetermined case)

$$\mathbf{z} = \mathbf{X}\mathbf{b}^{\mathrm{tru}} + \sigma \mathbf{e} \quad (n = 20, p = 30)$$

$\mathbf{X} \in \mathbb{R}^{20\times 30}$: generated by zero-mean Gaussian

$\|\mathbf{X}_{:i}\| = \sqrt{20}$ $(i = 1, 2, \ldots, 30)$ and $\mathbf{X}_{:2} = \mathbf{X}_{:3} = \mathbf{X}_{:4}$

$$\mathbf{b}^{\mathrm{tru}} = \frac{1}{\sqrt{30}}(0, 0, 0, 1, 1, 1, 0, 0, \ldots, 0)^t \in \mathbb{R}^{30}$$

$$T_j := \left(2\mathrm{prox}_{F_j} - \mathrm{Id}\right) \circ \left(2P_{\mathcal{M}_j} - \mathrm{Id}\right) : \mathbb{R}^{p+n+1} \to \mathbb{R}^{p+n+1} : \text{ Nonexpansive}$$

K-M algorithm with Douglas-Rachford Operator

$$(\mathbf{b}_{k+1}, \mathbf{c}_{k+1}) := (1 - \alpha_k)(\mathbf{b}_k, \mathbf{c}_k) + \alpha_k T_j(\mathbf{b}_k, \mathbf{c}_k) \quad (\alpha_k = 1.95/2)$$

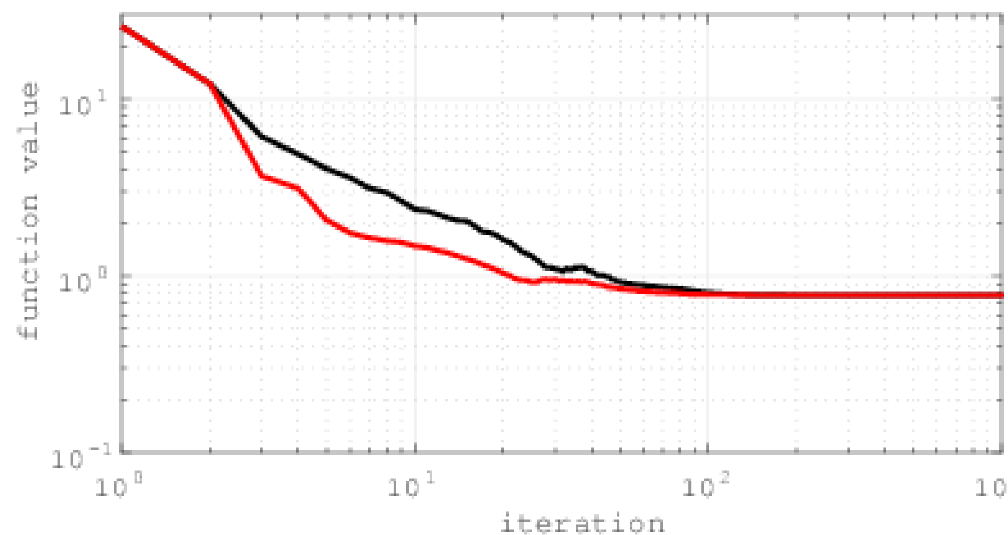
Hierarchical TREX for Promoting Flatness

$$(\mathbf{b}_{k+1}, \mathbf{c}_{k+1}) := T_j(\mathbf{b}_k, \mathbf{c}_k) - \lambda_{k+1} \nabla \Psi\left(T_j(\mathbf{b}_k, \mathbf{c}_k)\right) \quad (\lambda_k = 1/k)$$

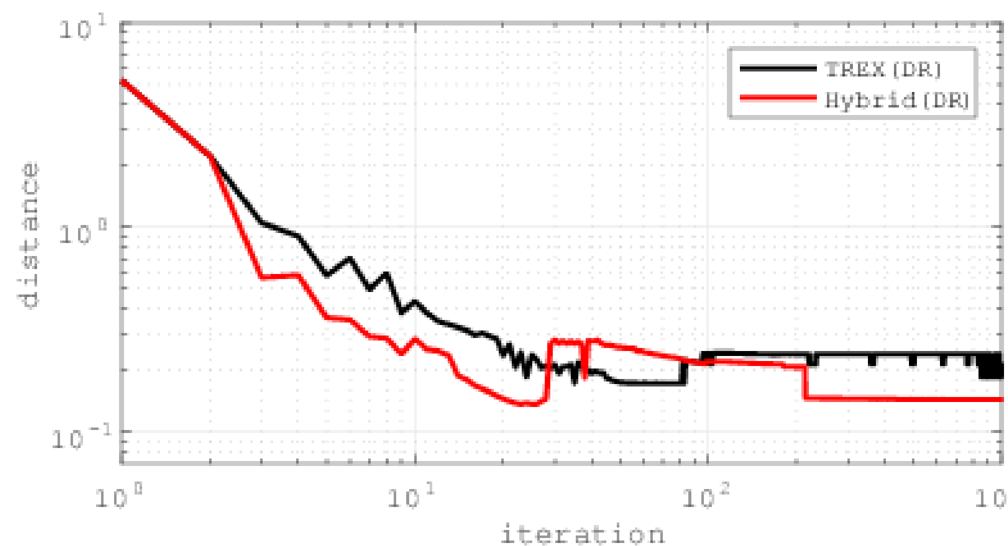
Numerical Performance

$$\mathrm{SNR} = 10\log_{10}\frac{\|X\mathbf{b}^{\mathrm{tru}}\|^2}{\|\sigma\mathbf{e}\|^2} = 20(\mathrm{dB})$$

$$\min_{1\leq j\leq 60}\{g_j(\mathbf{M}_j\mathbf{b}) + \|\mathbf{b}\|_1\}$$



$$\|\mathbf{b} - \mathbf{b}^{\mathrm{tru}}\|$$



Conclusion

1. We introduced a simple strategy for Hierarchical Convex Optimization which can enhance further existing proximal splitting algorithms without losing their optimality.

2. The proposed strategies are based on destined mariage: Proximal splitting operators + Hybrid steepest descent method.

3. We have demonstrated an application to Hierarchical Enhancement of Lasso estimator.

References (Hierarchical convex optimization)

[Yamada, Yukawa, Yamagishi 2011]

I. Yamada, M. Yukawa, M. Yamagishi, “Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings,“ In: Fixed-point algorithms for inverse problems in science and engineering, pp.343-388, Springer, 2011.

[Yamagishi, Yamada 2017]

M. Yamagishi, I. Yamada, “Nonexpansiveness of Linearlized Augmented Lagrangian operator for hierarchical convex optimization," Inverse Problems, 2017.