Recovery Guarantees for Low Complexity Models

Samuel Vaiter

CNRS & Université de Bourgogne

Joint work with J. Fadili & G. Peyré

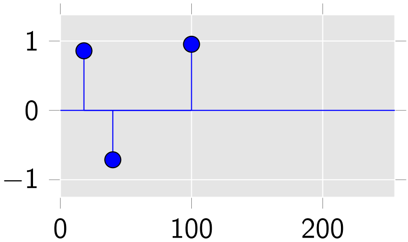
January 19, 2017

Aachen

Linear Inverse Problem

Recover data x_0 from observations y

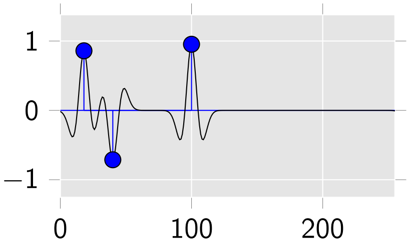
$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



Linear Inverse Problem

Recover data x_0 from observations y

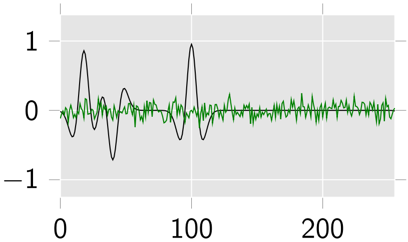
$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



Linear Inverse Problem

Recover data x_0 from observations y

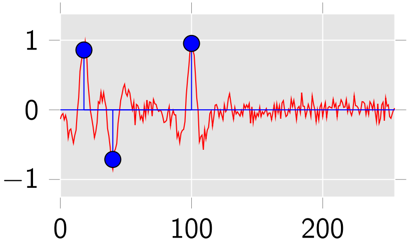
$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



Linear Inverse Problem

Recover data x_0 from observations y

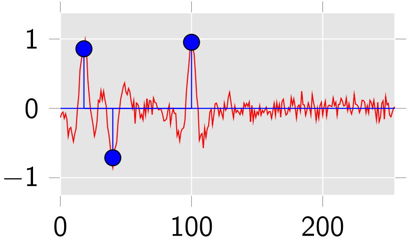
$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



Linear Inverse Problem

Recover data x_0 from observations y

$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



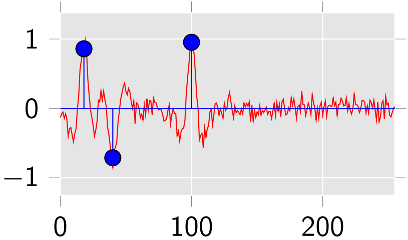
This talk:

- finite dimensional setting $\to$ $x_0 \in \mathbb{R}^n$, $y \in \mathbb{R}^q$
- No (explicit) assumption on the distribution of the noise w
- Φ is the *linear* measurement/degradation operator

Linear Inverse Problem

Recover data x_0 from observations y

$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



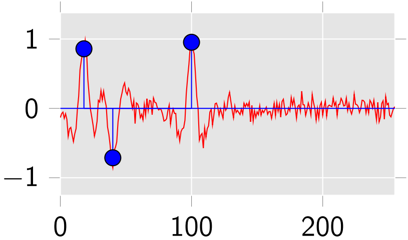
Here $\Phi x_0 = k \star x_0$. In Fourier domain,

$$\hat{x_0} = \frac{\hat{y}}{\hat{k}} - \frac{\hat{w}}{\hat{k}}$$

Linear Inverse Problem

Recover data x_0 from observations y

$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



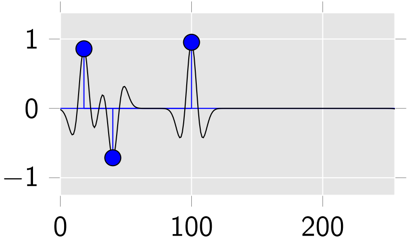
Here $\Phi x_0 = k \star x_0$. In Fourier domain,

$$\hat{x_0} = \frac{\hat{y}}{\hat{k}} - \frac{\hat{w}}{\hat{k}}$$

Linear Inverse Problem

Recover data x_0 from observations y

$$y = \underbrace{\Phi x_0}_{\mu_0}$$



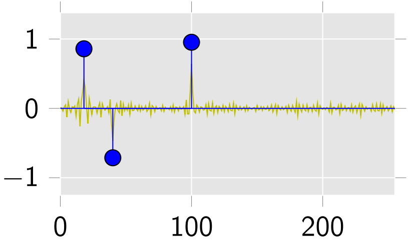
Here $\Phi x_0 = k \star x_0$. In Fourier domain,

$$\hat{x_0} = \frac{\hat{y}}{\hat{k}}$$

Linear Inverse Problem

Recover data x_0 from observations y

$$y = \underbrace{\Phi x_0}_{\mu_0}$$



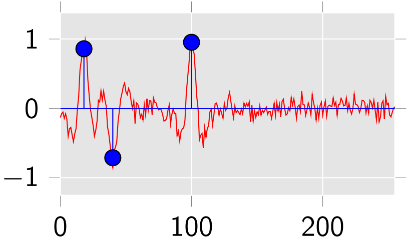
Here $\Phi x_0 = k \star x_0$. In Fourier domain,

$$\hat{x_0} = \frac{\hat{y}}{\hat{k}}$$

Linear Inverse Problem

Recover data x_0 from observations y

$$y = \underbrace{\Phi x_0}_{\mu_0} + w$$



Summary: the inverse problem of recovering x_0 from y is ill-posed

Regularization

As we just saw (in a different language), the optimization problem

$$\underset{x}{\operatorname{argmin}} \|y - \Phi x\|_2^2$$

leads to an unstable solution.

Regularization

As we just saw (in a different language), the optimization problem

$$\underset{x}{\operatorname{argmin}} \|y - \Phi x\|_2^2$$

leads to an unstable solution.

Could we cast an another optimization scheme with better properties ?

Regularization

As we just saw (in a different language), the optimization problem

$$\underset{x}{\operatorname{argmin}} \|y - \Phi x\|_2^2$$

leads to an unstable solution.

Could we cast an another optimization scheme with better properties ?
$\rightarrow$ idea of regularization.

Regularization

$$\underset{x}{\text{Argmin}} \; \|y - \Phi x\|_2^2 + \lambda J(x)$$

Regularization

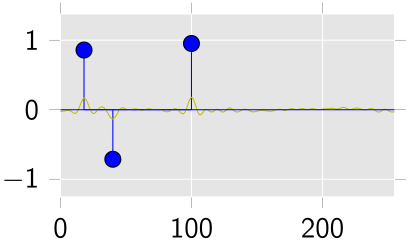
$$\underset{x}{\text{Argmin}} \; \|y - \Phi x\|_2^2 + \lambda \|x\|_2^2$$

Natural idea: a physical signal has a (relatively) low energy

Regularization

$$\underset{x}{\text{Argmin}}\ \|y - \Phi x\|_2^2 + \lambda \|x\|_2^2$$

Natural idea: a physical signal has a (relatively) low energy



Not very satisfying

Regularization

$$\underset{x}{\text{Argmin}} \; \|y - \Phi x\|_2^2 + \lambda J(x)$$

But wait, x_0 is a familly of spikes, why not just count them?

Regularization

$$\underset{x}{\text{Argmin}}\ \|y - \Phi x\|_2^2 + \lambda |\operatorname{supp}(x)|$$

But wait, x_0 is a familly of spikes, why not just count them?
$\rightarrow$ sparsity

Regularization

$$\underset{x}{\text{Argmin}}\ \|y - \Phi x\|_2^2 + \lambda |\operatorname{supp}(x)|$$

But wait, x_0 is a familly of spikes, why not just count them?
$\rightarrow$ sparsity
Issues: $x \mapsto |\operatorname{supp}(x)|$ is non-differentiable, non-convex

Regularization

$$\underset{x}{\text{Argmin}}\ \|y - \Phi x\|_2^2 + \lambda |\operatorname{supp}(x)|$$

But wait, x_0 is a familly of spikes, why not just count them?
$\rightarrow$ sparsity
Issues: $x \mapsto |\operatorname{supp}(x)|$ is non-differentiable, non-convex
Two alternatives:

- Use a greedy solver
- Embrace the power of convex relaxation

Regularization

$$\underset{x}{\text{Argmin}}\ \|y - \Phi x\|_2^2 + \lambda \|x\|_1$$

But wait, x_0 is a familly of spikes, why not just count them?
$\rightarrow$ sparsity
Issues: $x \mapsto |\,\text{supp}(x)|$ is non-differentiable, non-convex
Two alternatives:

- Use a greedy solver
- Embrace the power of convex relaxation

Regularization

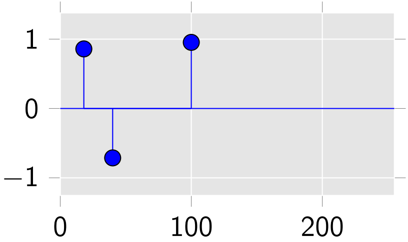
$$\underset{x}{\text{Argmin}} \; \|y - \Phi x\|_2^2 + \lambda \|x\|_1$$

But wait, x_0 is a familly of spikes, why not just count them?
$\to$ sparsity
Issues: $x \mapsto |\operatorname{supp}(x)|$ is non-differentiable, non-convex
Two alternatives:

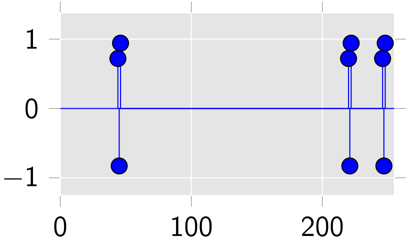
- Use a greedy solver
- Embrace the power of convex relaxation

The connection between $|\operatorname{supp}(\cdot)|$ and $\|\cdot\|_1$ is known as *compressed sensing* at Aachen

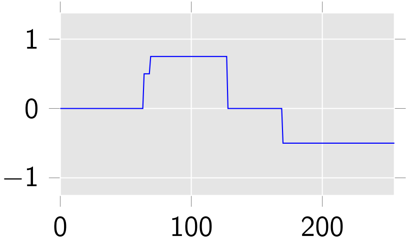
What If?



What If?



What If?



What If?

$$\begin{pmatrix} 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 2 & 4 & 2 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 2 & 4 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}$$

Main Assumption

x_0 lives in a low-dimensional submanifold of $\mathbb{R}^n$

Main Assumption

x_0 lives in a low-dimensional submanifold of $\mathbb{R}^n$

e.g. x_0 is sparse, block-sparse, piecewise constant, piecewise affine, low-rank, etc.

Main Assumption

x_0 lives in a low-dimensional submanifold of $\mathbb{R}^n$

e.g. x_0 is sparse, block-sparse, piecewise constant, piecewise affine, low-rank, etc.

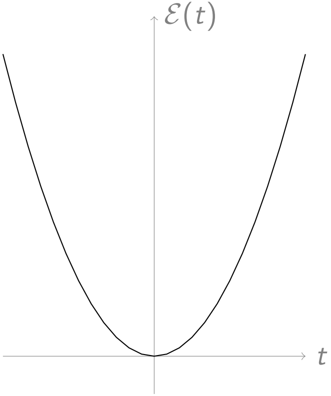
Our goal: encompass all these priors under a single (convex) umbrella.

Convex Analysis 101: Euler Equation

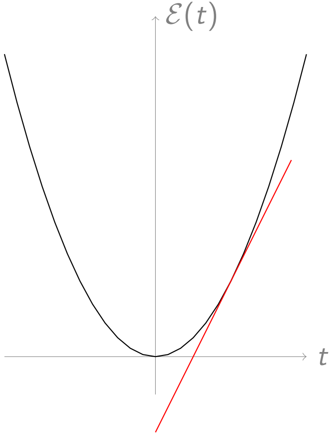
$\mathcal{E}$ convex + smooth

$$0 = \nabla \mathcal{E}(x^\star) \iff x^\star \in \operatorname*{Argmin}_x \mathcal{E}(x)$$

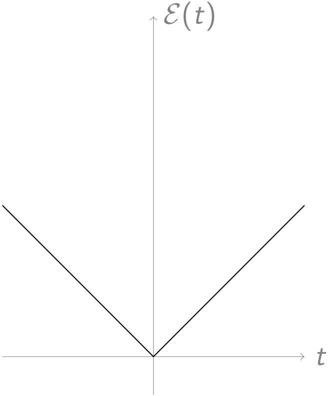
Convex Analysis 101: Subdifferential



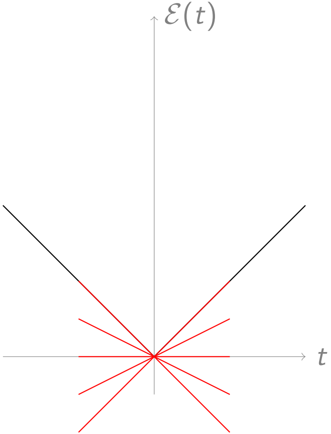
Convex Analysis 101: Subdifferential



Convex Analysis 101: Subdifferential



Convex Analysis 101: Subdifferential



$$\partial\mathcal{E}(t) = \left\{\eta :\ \mathcal{E}(t') \geqslant \mathcal{E}(t) + \langle \eta,\ t' - t\rangle\right\}$$

Convex Analysis 101: Euler Equation

$\mathcal{E}$ convex + smooth

$$0 = \nabla\mathcal{E}(x^\star) \Longleftrightarrow x^\star \in \underset{x}{\text{Argmin}}\ \mathcal{E}(x)$$

$\mathcal{E}$ convex

$$0 \in \partial\mathcal{E}(x^\star) \Longleftrightarrow x^\star \in \underset{x}{\text{Argmin}}\ \mathcal{E}(x)$$

Part I: ℓ^2-stability

In order to solve

$$y = \Phi x_0 + w$$

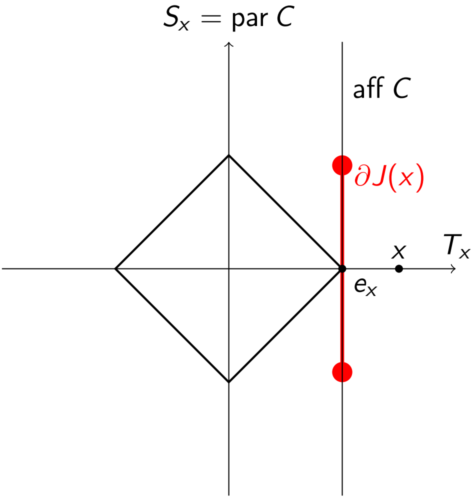
we consider for a convex function J, the optimization

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \; \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Goal

Provide an upper bound of the estimation error $\|x^\star_{y,\lambda} - x_0\|$

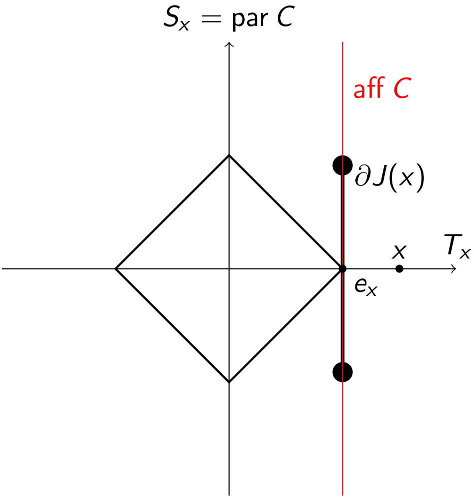
Notations



$C = \partial J(x)$: subdifferential of J at x
aff(C): affine hull of C
ri C: relative interior of C
par(C): subspace parallel to aff(C)

$$S_x = \text{par}\, \partial J(x),\ T_x = S_x^{\perp},\ e_x = \Pi_{T_x}(\partial J(x))$$

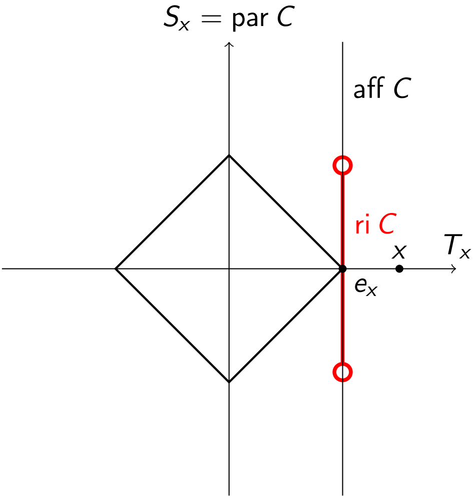
Notations



$C = \partial J(x)$: subdifferential of J at x
aff(C): affine hull of C
ri C: relative interior of C
par(C): subspace parallel to aff(C)

$$S_x = \text{par}\, \partial J(x),\ T_x = S_x^{\perp},\ e_x = \Pi_{T_x}(\partial J(x))$$

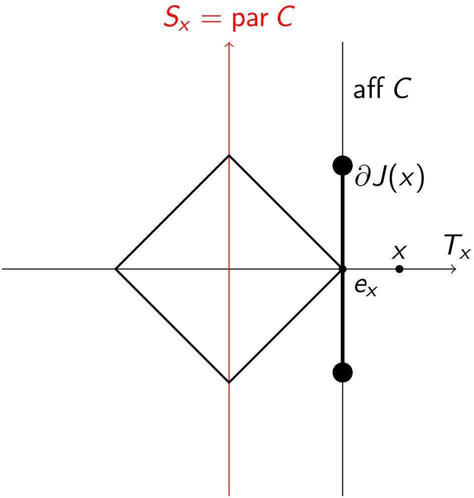
Notations



$C = \partial J(x)$: subdifferential of J at x
aff(C): affine hull of C
ri C: relative interior of C
par(C): subspace parallel to aff(C)

$$S_x = \text{par}\, \partial J(x),\ T_x = S_x^{\perp},\ e_x = \Pi_{T_x}(\partial J(x))$$

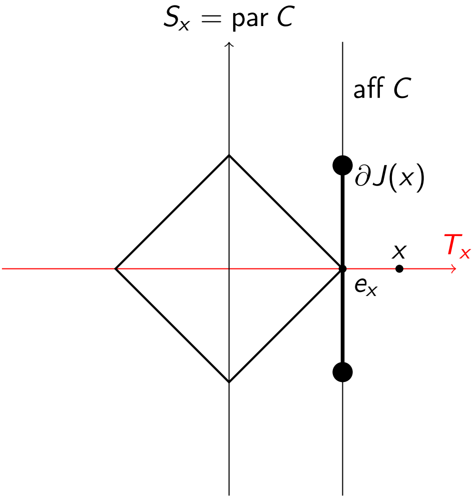
Notations



$C = \partial J(x)$: subdifferential of J at x
aff(C): affine hull of C
ri C: relative interior of C
par(C): subspace parallel to aff(C)

$$S_x = \text{par}\, \partial J(x),\ T_x = S_x^{\perp},\ e_x = \Pi_{T_x}(\partial J(x))$$

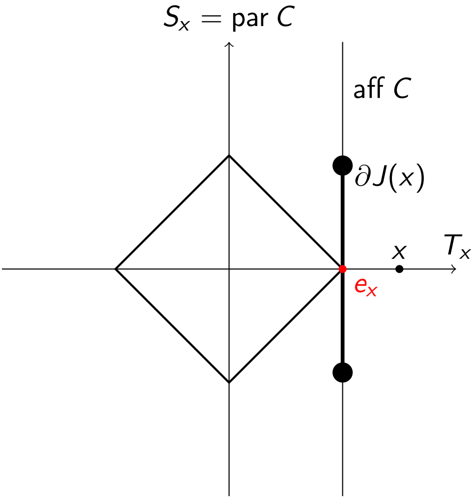
Notations



$C = \partial J(x)$: subdifferential of J at x
aff(C): affine hull of C
ri C: relative interior of C
par(C): subspace parallel to aff(C)

$$S_x = \mathrm{par}\, \partial J(x),\ T_x = S_x^{\perp},\ e_x = \Pi_{T_x}(\partial J(x))$$

Notations



$C = \partial J(x)$: subdifferential of J at x
aff(C): affine hull of C
ri C: relative interior of C
par(C): subspace parallel to aff(C)

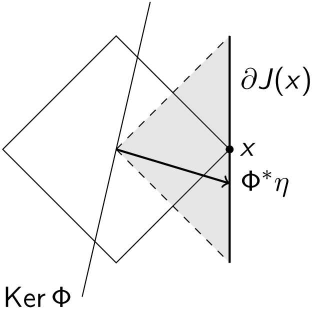
$$S_x = \text{par}\, \partial J(x),\ T_x = S_x^{\perp},\ e_x = \Pi_{T_x}(\partial J(x))$$

(Non-Degenerated) Source Condition

$$\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \partial J(x) \qquad (\mathrm{SC}_x)$$

(Non-Degenerated) Source Condition

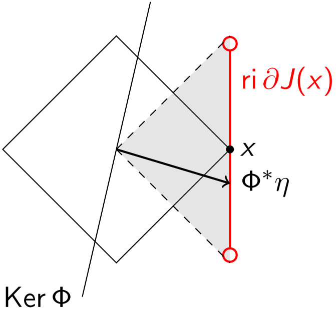
$$\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \partial J(x) \tag{SC$_x$}$$



$$\begin{aligned} & x \in \underset{u}{\text{Argmin}}\ J(u) \text{ s.t } \Phi x = \Phi u \\ \Leftrightarrow \quad & 0 \in \partial J(x) + \mathcal{N}_{\text{Ker}\,\Phi}(x) \\ \Leftrightarrow \quad & 0 \in \partial J(x) + \text{Im}\,\Phi^* \\ \Leftrightarrow \quad & \text{Im}\,\Phi^* \cap \partial J(x) \neq \emptyset \end{aligned}$$

(Non-Degenerated) Source Condition

$$\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \partial J(x) \qquad (\mathrm{SC}_x)$$



$$x \in \underset{u}{\text{Argmin}}\ J(u) \text{ s.t } \Phi x = \Phi u$$
$$\Leftrightarrow \quad 0 \in \partial J(x) + \mathcal{N}_{\text{Ker}\,\Phi}(x)$$
$$\Leftrightarrow \quad 0 \in \partial J(x) + \text{Im}\,\Phi^*$$
$$\Leftrightarrow \quad \text{Im}\,\Phi^* \cap \partial J(x) \neq \emptyset$$

Non-degenerated source condition

$$\boxed{\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \text{ri}\,\partial J(x)} \qquad (\widetilde{\mathrm{SC}}_x)$$

Restricted Injectivity

$$\boxed{\operatorname{Ker}\Phi \cap T = \{0\}} \qquad (\mathrm{INJ}_T)$$

Observe that if $y = \Phi x_0 + 0$ and $x_0 \in T$ (known). Then,

$$x_0 = \underset{\Phi x = y}{\operatorname{argmin}}\ J(x) \Leftrightarrow (\mathrm{INJ}_T) \text{ holds}$$

We proved a uniqueness result based on this remark (NSP-like, not covered today easy question!)

ℓ^2-stability

$$y = \underbrace{\Phi x_0}_{=y_0} + w$$

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \; \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Theorem

Assume $(\widetilde{\mathrm{SC}}_{x_0})$*, associated to a non-degenerate certificate* η*, and* (INJ_T) *hold. Choosing* $\lambda = c\|w\|_2$*,* $c > 0$*, for any minimizer* $x^\star_{y,\lambda}$ *of* $(\mathcal{P}_{y,\lambda})$

$$\|x^\star_{y,\lambda} - x^\star_{y_0,0^+}\|_2 \leqslant C(c, \Phi, \eta)\|w\|_2$$

ℓ^2-stability

$$y = \underbrace{\Phi x_0}_{=y_0} + w$$

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Theorem

Assume $(\widetilde{\mathrm{SC}}_{x_0})$*, associated to a non-degenerate certificate* η*, and* (INJ_T) *hold. Choosing* $\lambda = c\|w\|_2$*,* $c > 0$*, for any minimizer* $x^\star_{y,\lambda}$ *of* $(\mathcal{P}_{y,\lambda})$

$$\|x^\star_{y,\lambda} - x_0\|_2 = O(\|w\|_2)$$

Previous works:

- [Grasmair et al. 2010]: ℓ^1
- [Grasmair 2011]: $J(x^\star_{y,\lambda} - x_0) = O(\|w\|_2)$
- [Haltmeier 2012]: analysis-ℓ^1 with a frame

ℓ^2-stability

- $\|x^\star_{y,\lambda} - x^\star_{y_0,0^+}\|_2 \leqslant C(c,\Phi,\eta)\|w\|_2$ provides a *worst case* bound

ℓ^2-stability

- $\|x^\star_{y,\lambda} - x^\star_{y_0,0^+}\|_2 \leqslant C(c, \Phi, \eta)\|w\|_2$ provides a *worst case* bound
- The use of ℓ^2-norm is not central. Other data fidelity term can be considered

ℓ^2-stability

- $\|x^\star_{y,\lambda} - x^\star_{y_0,0^+}\|_2 \leqslant C(c, \Phi, \eta)\|w\|_2$ provides a *worst case* bound
- The use of ℓ^2-norm is not central. Other data fidelity term can be considered
- A similar analysis can be performed for the constrained case, i.e.

$$\underset{x\in\mathbb{R}^n}{\text{Argmin}}\ J(x) \quad \text{s.t} \quad \|y - \Phi x\|_2 \leqslant \varepsilon$$

ℓ^2-stability

- $\|x^\star_{y,\lambda} - x^\star_{y_0,0^+}\|_2 \leqslant C(c, \Phi, \eta)\|w\|_2$ provides a *worst case* bound
- The use of ℓ^2-norm is not central. Other data fidelity term can be considered
- A similar analysis can be performed for the constrained case, i.e.

$$\underset{x \in \mathbb{R}^n}{\text{Argmin}}\ J(x) \quad \text{s.t} \quad \|y - \Phi x\|_2 \leqslant \varepsilon$$

ℓ^2-stability

- $\|x^\star_{y,\lambda} - x^\star_{y_0,0^+}\|_2 \leqslant C(c, \Phi, \eta)\|w\|_2$ provides a *worst case* bound
- The use of ℓ^2-norm is not central. Other data fidelity term can be considered
- A similar analysis can be performed for the constrained case, i.e.

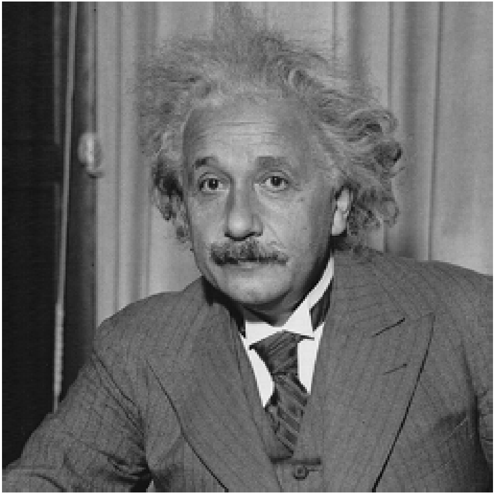
$$\underset{x \in \mathbb{R}^n}{\text{Argmin}}\ J(x) \quad \text{s.t} \quad \|y - \Phi x\|_2 \leqslant \varepsilon$$

Connection to compressed sensing

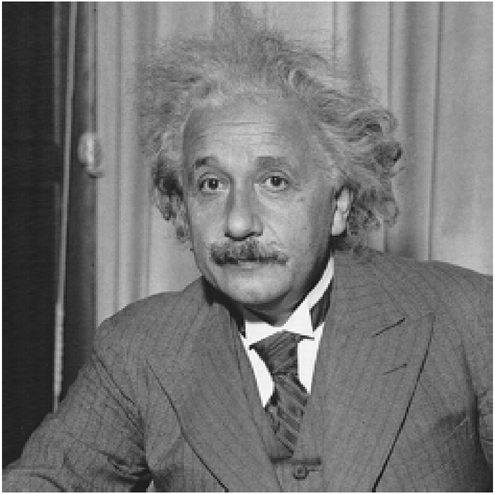
Proposition

Assume $J = \|\cdot\|_1$, $\Phi_{ij} \sim_{iid} \mathcal{N}(0,1)$ *and* $s = \|x_0\|_0$.
If $q > 2s\log(n/s) + 7/5s$, *then* $(\widetilde{\mathrm{SC}}_{x_0})$ *and* (INJ_T) *hold.*

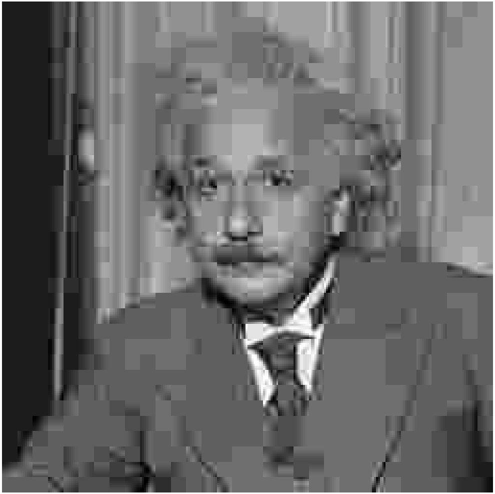
Should I be Happy ?



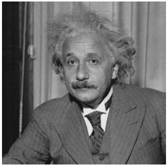
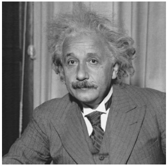
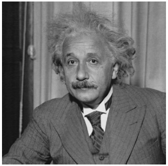
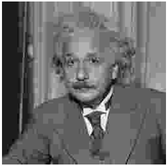
Should I be Happy ?



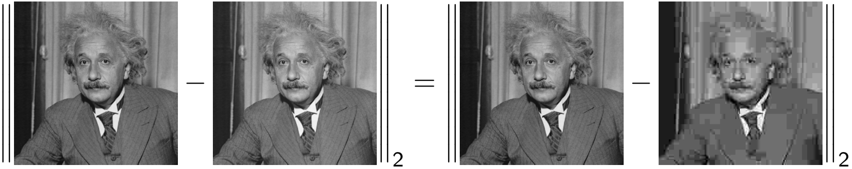
Should I be Happy ?



Should I be Happy ?

$\Big\|$  $-$  $\Big\|_2 =$ $\Big\|$  $-$  $\Big\|_2$

Should I be Happy ?



We need to use our main assumption

x_0 lives in a low-dimensional submanifold of $\mathbb{R}^n$

Part II: Definition of a Model

In order to solve

$$y = \Phi x_0 + w$$

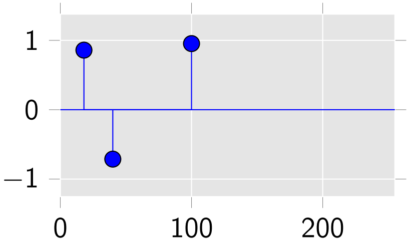
we consider for a convex function J, the optimization

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}} \ \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Goal

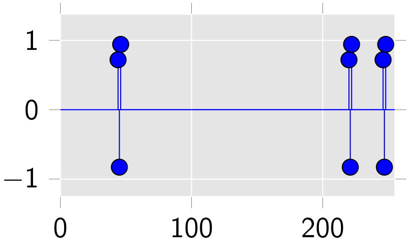
Connect a convex function J to a signal model (geometric / combinatorial) $\mathcal{M}$

Some Priors



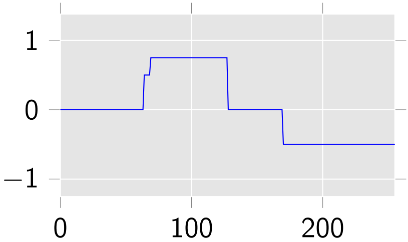
Type	Sparse signal
Complexity measure	$\|\cdot\|_0 = \|\operatorname{supp}(\cdot)\|$
Natural model:	$\{z : \operatorname{supp}(z) = \operatorname{supp}(x)\}$
Convex candidate:	$\|\cdot\|_1$

Some Priors



Type	Block-sparse signal
Complexity measure	$\lvert \mathrm{supp}_{\mathcal{B}}(\cdot) \rvert$
Natural model:	$\{z : \mathrm{supp}_{\mathcal{B}}(z) = \mathrm{supp}_{\mathcal{B}}(x)\}$
Convex candidate:	$\lVert \cdot \rVert_{\mathcal{B}}$

Some Priors



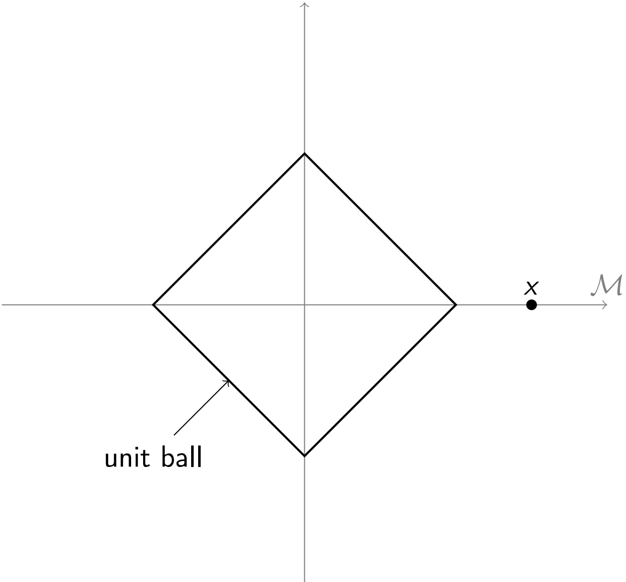
Type	Piecewise constant signal
Complexity measure	$\|\nabla \cdot\|_0 = \lvert \mathrm{supp}(\nabla \cdot) \rvert$
Natural model:	$\{z : \mathrm{supp}(\nabla z) = \mathrm{supp}(\nabla x)\}$
Convex candidate:	$\|\nabla \cdot\|_1$

Some Priors

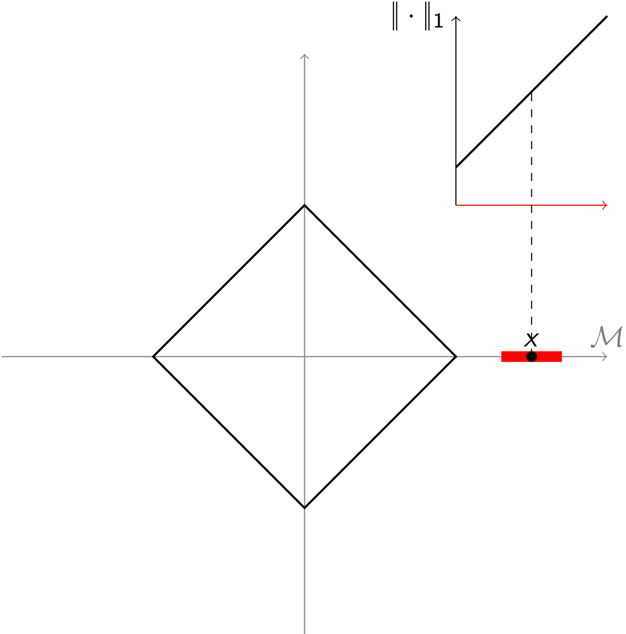
$$\begin{pmatrix} 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 2 & 4 & 2 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 2 & 4 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}$$

Type	Low rank matrix
Complexity measure	$\|\sigma(\cdot)\|_0 = \mathrm{rank}(\cdot)$
Natural model:	$\{z : \mathrm{rank}(z) = \mathrm{rank}(x)\}$
Convex candidate:	$\|\cdot\|_*$

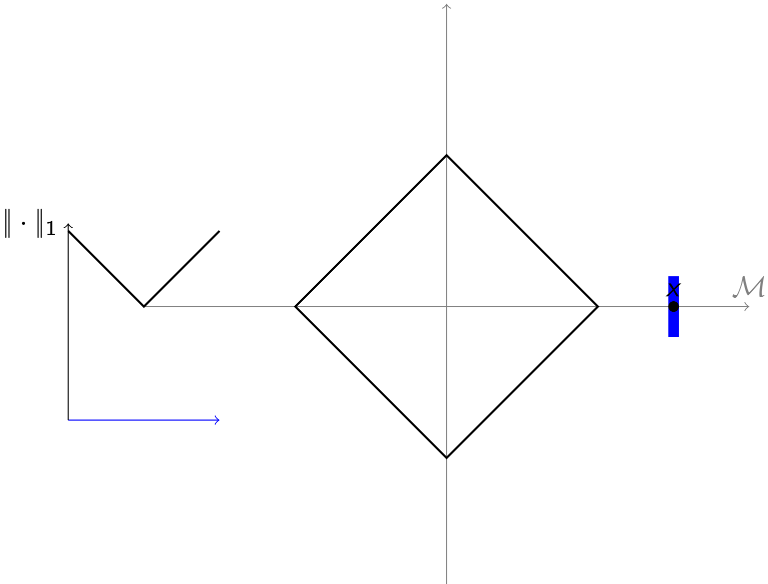
Back to ℓ^1



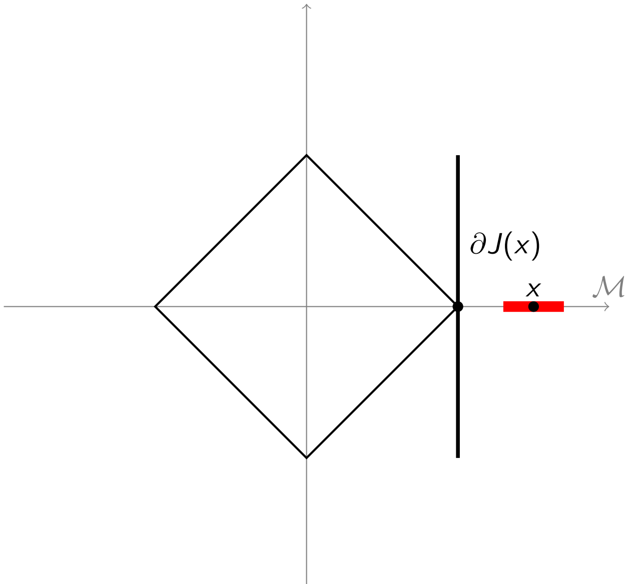
Back to ℓ^1



Back to ℓ^1



Back to ℓ^1



Partial Smoothness

1. $(\| \cdot \|_1)_{|\mathcal{M}}$ is affine around x
2. For every $h \in \mathcal{M}^{\perp}$, $t \mapsto \|x + th\|_1$ is not differentiable at 0
3. $(\partial \| \cdot \|_1)_{|\mathcal{M}}$ is constant around x relatively to $\mathcal{M}$

Partial Smoothness

1. $(\|\cdot\|_1)_{|\mathcal{M}}$ is C^2 around x
2. For every $h \in \mathcal{M}^\perp$, $t \mapsto \|x + th\|_1$ is not differentiable at 0
3. $(\partial\|\cdot\|_1)_{|\mathcal{M}}$ is continuous around x relatively to $\mathcal{M}$

Partial Smoothness

J is said to be partly smooth relatively to a C^2-manifold $\mathcal{M}$ at x if:

1. $J_{|\mathcal{M}}$ is C^2 around x
2. For every $h \in (\mathcal{T}_x\mathcal{M})^\perp$, $t \mapsto \|x + th\|_1$ is not differentiable at 0
3. $(\partial J)_{|\mathcal{M}}$ is continuous around x relatively to $\mathcal{M}$

Partial Smoothness

J is said to be partly smooth relatively to a C^2-manifold $\mathcal{M}$ at x if:

1. $J_{|\mathcal{M}}$ is C^2 around x
2. For every $h \in (\mathcal{T}_x\mathcal{M})^\perp$, $t \mapsto \|x + th\|_1$ is not differentiable at 0
3. $(\partial J)_{|\mathcal{M}}$ is continuous around x relatively to $\mathcal{M}$

Notation: $J \in \mathrm{PS}_x(\mathcal{M})$

Introduced by [Lewis 2002] following [Lemaréchal et al. 2000]

Proposition (Locally uniquely defined)

If $J \in \mathrm{PS}_x(\mathcal{M})$ *and* $J \in \mathrm{PS}_x(\mathcal{M}')$ *then*

$$\mathcal{M} \equiv^x \mathcal{M}'$$

Partial Smoothness

J is said to be partly smooth relatively to a C^2-manifold $\mathcal{M}$ at x if:

1. $J_{|\mathcal{M}}$ is C^2 around x
2. $\mathcal{T}_x\mathcal{M} = T_x \quad (= \mathrm{par}\, \partial J(x))$
3. $(\partial J)_{|\mathcal{M}}$ is continuous around x relatively to $\mathcal{M}$

Notation: $J \in \mathrm{PS}_x(\mathcal{M})$

Introduced by [Lewis 2002] following [Lemaréchal et al. 2000]

Proposition (Locally uniquely defined)

If $J \in \mathrm{PS}_x(\mathcal{M})$ *and* $J \in \mathrm{PS}_x(\mathcal{M}')$ *then*

$$\mathcal{M} \equiv^x \mathcal{M}'$$

Model Manifold

J	$\mathcal{M}$	
$\|\cdot\|_1$	$\{z : \operatorname{supp}(z) \subseteq \operatorname{supp}(x)\}$	same support
$\|\cdot\|_{\mathcal{B}}$	$\{z : \operatorname{supp}_{\mathcal{B}}(z) \subseteq \operatorname{supp}_{\mathcal{B}}(x)\}$	same block-support
$\|\nabla\cdot\|_1$	$\{z : \operatorname{supp}(\nabla z) \subseteq \operatorname{supp}(\nabla x)\}$	same jump set
$\|\cdot\|_*$	$\{z : \operatorname{rank} z = \operatorname{rank} x\}$	same rank
$\|\cdot\|_\infty$	$\{z : z_I \in \mathbb{R}\operatorname{sign}(x_I)\}$	same saturation[1]

[1] $I = \{i : |x_i| = \|x\|_\infty\}$

Calculus Rules

under mild transversality condition:

Proposition (Lewis 2002, Daniilidis et al. 2014)

- *If J is C^2 around x then $J \in \mathrm{PS}_x(\mathbb{R}^n)$*
- *If $J \in \mathrm{PS}_x(\mathcal{M})$ and $J' \in \mathrm{PS}_x(\mathcal{M}')$ then*

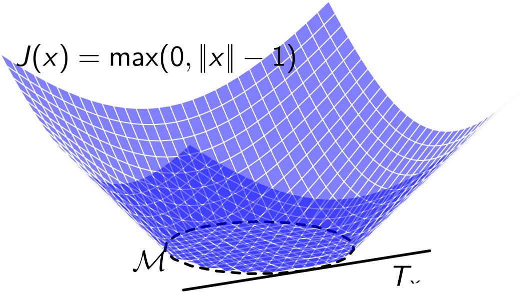
$$J + J' \in \mathrm{PS}_x(\mathcal{M} \cap \mathcal{M}')$$

- *If A is a linear operator and $J \in \mathrm{PS}_{Ax}(\mathcal{M}^0)$ then*

$$J \circ A \in \mathrm{PS}_x(\mathcal{M}) \quad \textit{where} \quad \mathcal{M} = \left\{ z : \ Au \in \mathcal{M}^0 \right\}$$

- *Spectral lift*

Fun Example



Part III: Model Stability

In order to solve

$$y = \Phi x_0 + w$$

we consider for a partly smooth function J, the optimization

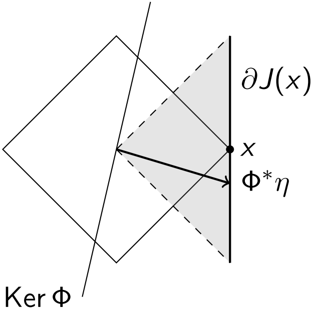
$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \ \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Goal

How to assess that $\mathcal{M}(x^\star_{y,\lambda}) = \mathcal{M}(x_0)$ and also $\|x^\star_{y,\lambda} - x_0\|$ small enough ?

(Non-Degenerated) Source Condition

$$\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \partial J(x) \qquad (\mathrm{SC}_x)$$



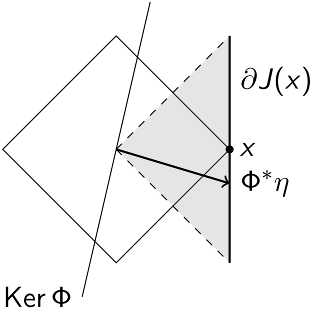
$$x \in \underset{u}{\text{Argmin}}\ J(u) \text{ s.t } \Phi x = \Phi u$$
$$\Leftrightarrow \quad 0 \in \partial J(x) + \mathcal{N}_{\text{Ker}\,\Phi}(x)$$
$$\Leftrightarrow \quad 0 \in \partial J(x) + \text{Im}\,\Phi^*$$
$$\Leftrightarrow \quad \text{Im}\,\Phi^* \cap \partial J(x) \neq \emptyset$$

Non-degenerated source condition

$$\boxed{\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \text{ri}\,\partial J(x)} \qquad (\widetilde{\mathrm{SC}}_x)$$

(Non-Degenerated) Source Condition

$$\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \partial J(x) \qquad (\mathrm{SC}_x)$$



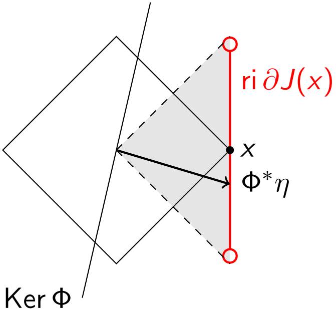
$$x \in \underset{u}{\operatorname{Argmin}}\ J(u) \text{ s.t } \Phi x = \Phi u$$
$$\Leftrightarrow \quad 0 \in \partial J(x) + \mathcal{N}_{\operatorname{Ker}\Phi}(x)$$
$$\Leftrightarrow \quad 0 \in \partial J(x) + \operatorname{Im}\Phi^*$$
$$\Leftrightarrow \quad \operatorname{Im}\Phi^* \cap \partial J(x) \neq \emptyset$$

Non-degenerated source condition

$$\boxed{\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \operatorname{ri} \partial J(x)} \qquad (\widetilde{\mathrm{SC}}_x)$$

(Non-Degenerated) Source Condition

$$\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \partial J(x) \qquad (\mathrm{SC}_x)$$



$$x \in \underset{u}{\text{Argmin}}\ J(u) \text{ s.t } \Phi x = \Phi u$$

$$\Leftrightarrow \quad 0 \in \partial J(x) + \mathcal{N}_{\text{Ker}\,\Phi}(x)$$

$$\Leftrightarrow \quad 0 \in \partial J(x) + \text{Im}\,\Phi^*$$

$$\Leftrightarrow \quad \text{Im}\,\Phi^* \cap \partial J(x) \neq \emptyset$$

Non-degenerated source condition

$$\boxed{\exists \eta \in \mathbb{R}^q \quad \text{s.t} \quad \Phi^* \eta \in \text{ri}\,\partial J(x)} \qquad (\widetilde{\mathrm{SC}}_x)$$

Minimal Norm (Pre-)Certificate

How to exhibit a certificate $\eta \in \mathbb{R}^q$ s.t $\Phi^*\eta \in \operatorname{ri} \partial J(x)$?

Could be a hard problem.

Minimal Norm (Pre-)Certificate

How to exhibit a certificate $\eta \in \mathbb{R}^q$ s.t $\Phi^* \eta \in \operatorname{ri} \partial J(x)$?

Could be a hard problem.

Minimal-norm certificate

$$\eta_0 = \underset{\eta \in \mathbb{R}^q}{\operatorname{argmin}} \|\eta\|_2 \text{ s.t. } \Phi^* \eta \in \partial J(x)$$

Minimal Norm (Pre-)Certificate

How to exhibit a certificate $\eta \in \mathbb{R}^q$ s.t $\Phi^* \eta \in \mathrm{ri}\, \partial J(x)$?

Could be a hard problem.

Minimal-norm certificate

$$\eta_0 = \underset{\eta \in \mathbb{R}^q}{\operatorname{argmin}} \|\eta\|_2 \text{ s.t. } \Phi^* \eta \in \partial J(x)$$

Linearized pre-certificate ($T = T_x$)

$$\eta_F = \underset{\eta \in \mathbb{R}^q}{\operatorname{argmin}} \|\eta\|_2 \text{ s.t. } \Phi^* \eta \in \mathrm{aff}\, \partial J(x)$$

Minimal Norm (Pre-)Certificate

How to exhibit a certificate $\eta \in \mathbb{R}^q$ s.t $\Phi^* \eta \in \operatorname{ri} \partial J(x)$?

Could be a hard problem.

Minimal-norm certificate

$$\eta_0 = \underset{\eta \in \mathbb{R}^q}{\operatorname{argmin}} \|\eta\|_2 \text{ s.t. } \Phi^* \eta \in \partial J(x)$$

Linearized pre-certificate ($T = T_x$)

$$\eta_F = \underset{\eta \in \mathbb{R}^q}{\operatorname{argmin}} \|\eta\|_2 \text{ s.t. } \Phi^* \eta \in \operatorname{aff} \partial J(x)$$

Proposition

If (INJ_T) *holds, then*

1. η_F *is well-defined*
2. $\eta_F = (\Phi \Pi_T)^{+,*} e_x$
3. $\Phi^* \eta_F \in \operatorname{ri} \partial J(x) \Rightarrow \eta_0 = \eta_F$
4. $\Phi^* \eta_F \in \operatorname{ri} \partial J(x) \Rightarrow x$ *unique solution of* $(\mathcal{P}_{y_0, 0^+})$

Model Stability

$$x^\star_{y,\lambda} \in \underset{x\in\mathbb{R}^n}{\text{Argmin}} \ \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Theorem

Assume that $J \in \mathrm{PS}_{x_0}(\mathcal{M})$, $(\mathrm{INJ}_{T_{x_0}})$ and $\Phi^*\eta_F \in \mathrm{ri}\,\partial J(x_0)$ hold. Then, there exist (C, C') such that if $\|w\|_2 \leqslant C$ and $\lambda = C'\|w\|_2$, the solution $x^\star_{y,\lambda}$ of $(\mathcal{P}_{y,\lambda})$ is unique,

$$x^\star_{y,\lambda} \in \mathcal{M} \quad \textit{and} \quad \|x^\star_{y,\lambda} - x_0\|_2 = O(\|w\|_2)$$

Model Stability

$$x^\star_{y,\lambda} \in \underset{x\in\mathbb{R}^n}{\text{Argmin}} \; \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Theorem

Assume that $J \in \mathrm{PS}_{x_0}(\mathcal{M})$, $(\mathrm{INJ}_{T_{x_0}})$ and $\Phi^*\eta_F \in \operatorname{ri}\partial J(x_0)$ hold. Then, there exist (C, C') such that if $\|w\|_2 \leqslant C$ and $\lambda = C'\|w\|_2$, the solution $x^\star_{y,\lambda}$ of $(\mathcal{P}_{y,\lambda})$ is unique,

$$x^\star_{y,\lambda} \in \mathcal{M} \quad \textit{and} \quad \|x^\star_{y,\lambda} - x_0\|_2 = O(\|w\|_2)$$

Previous works: [**F**uchs 2004] ℓ^1, [Bach 2008] $\ell^1 - \ell^2$, [V. et al. 2012] analysis-ℓ^1

For most J, $x^\star_{y,\lambda} \in \mathcal{M} \Rightarrow \mathcal{M}(x^\star_{y,\lambda}) = \mathcal{M}$

Almost sharp, i.e. $\Phi^*\eta_F \notin \partial J(x_0) \Rightarrow$ no model stability

Gaussian Measurements

Previous theorem: two non trivial hypotheses $(\mathrm{INJ}_{T_{x_0}})$ and $\Phi^* \eta_F \in \mathrm{ri}\, \partial J(x_0)$

Proposition

Suppose $\Phi_{ij} \sim_{iid} \mathcal{N}(0,1)$. *If, either*

1. $J = \|\cdot\|_1$, $s = \|x_0\|_0$ *and* $q > 2\beta s \log n + s$ *for some* $\beta > 1$
2. $J = \|\cdot\|_*$, $r = \mathrm{rank}(x_0)$ *and* $q > \beta r(6\sqrt{n} - 5r)$ *for some* $\beta > 1$

then $(\mathrm{INJ}_{T_{x_0}})$ *and* $\Phi^* \eta_F \in \mathrm{ri}\, \partial J(x_0)$

ℓ^2-stability: $2s \log n/s$ measures VS model stability: $2s \log n$

Part IV: Algorithmic Implication

In order to solve

$$y = \Phi x_0 + w$$

we consider for a partly smooth function J, the optimization

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \; \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Goal

How to assess that an algorithm provides the good model in finite time?

Algorithm and Non-smoothness

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}}\ \mathcal{E}(x) = F(x) + \lambda J(x)$$

Gradient descent

$$x^{(k+1)} = x^{(k)} - \rho \nabla \mathcal{E}(x^{(k)})$$

Algorithm and Non-smoothness

$$x^\star \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}}\ \mathcal{E}(x) = F(x) + \lambda J(x)$$

Gradient descent

$$x^{(k+1)} = x^{(k)} - \rho \nabla \mathcal{E}(x^{(k)})$$

But J is not smooth. Several strategies:

- Smoothing of J
- Interior point method
- Subgradient descent
- Proximal methods

Algorithm and Non-smoothness

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}}\ \mathcal{E}(x) = F(x) + \lambda J(x)$$

Gradient descent

$$x^{(k+1)} = x^{(k)} - \rho \nabla \mathcal{E}(x^{(k)})$$

But J is not smooth. Several strategies:

- Smoothing of J
- Interior point method
- Subgradient descent
- Proximal methods

Life is Smooth: Moreau–Yosida

Infimal convolution

$$(f \Box g)(x) = \inf_v f(x) + g(v - x)$$

Moreau-Yosida regularization

$$\mathrm{Mor}[f] = f \Box (1/2)\| \cdot \|^2$$

Life is Smooth: Moreau–Yosida

Infimal convolution

$$(f \Box g)(x) = \inf_v f(x) + g(v - x)$$

Moreau-Yosida regularization

$$\mathrm{Mor}[f] = f \Box (1/2)\| \cdot \|^2$$

For any convex function f (not smooth, not full-domain)

- $\mathrm{dom}\,\mathrm{Mor}[f] = \mathbb{R}^n$
- $\mathrm{Mor}[f]$ is continuously differentiable
- $\mathrm{argmin}\,\mathrm{Mor}[f] = \mathrm{argmin}\, f$

Proximity Operator

Proximity operator $\equiv$ unique argument of Moreau infimum

$$\mathrm{Prox}_f(v) = \underset{x \in \mathbb{R}^n}{\mathrm{argmin}}\ f(x) + \frac{1}{2}\|x - v\|_2^2$$

Smooth interpretation: *implicit* gradient step

$$\mathrm{Prox}_f(x) = x - \nabla \mathrm{Mor}[f](x)$$

Proximity $\approx$ Generalized Projection

Indicator function

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition (Proximity $\equiv$ Projection)

If $\mathcal{C}$ is a convex set, then

$$\mathrm{Prox}_{\iota_{\mathcal{C}}} = \Pi_{\mathcal{C}}$$

$$\begin{aligned} \mathrm{Prox}_{\iota_{\mathcal{C}}}(v) &= \operatorname*{argmin}_{x} \ \mathrm{Prox}_{\iota_{\mathcal{C}}}(v) + \frac{1}{2}\|x - v\|^2 \\ &= \operatorname*{argmin}_{x \in \mathcal{C}} \frac{1}{2}\|x - v\|^2 = \Pi_{\mathcal{C}}(v) \end{aligned}$$

Subdifferential and Proximity Operator

Proposition

$$p = \mathrm{Prox}_f(v) \Leftrightarrow v - p \in \partial f(p)$$

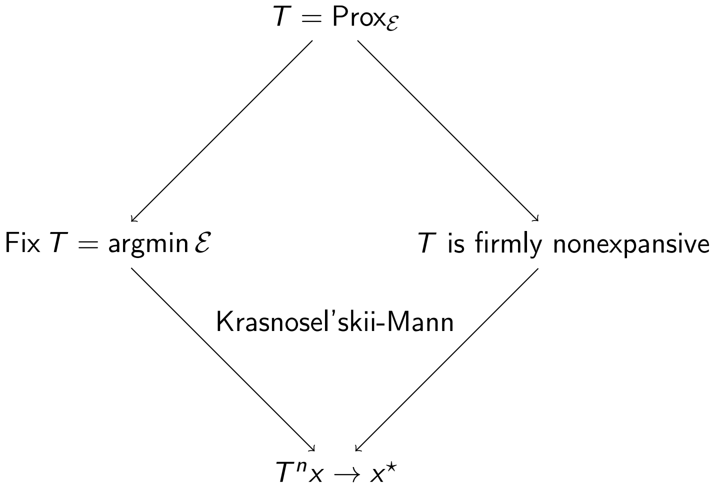
Resolvant of the subdifferential (as a notation)

$$\mathrm{Prox}_f(v) = (\mathrm{Id} + \partial f)^{-1}(v)$$

Theorem

$$\mathrm{Fix}\,\mathrm{Prox}_f = \operatorname{argmin} f$$

Proximal Fixed Point



Firmly nonexpansive

$$\|\mathrm{Prox}_{\mathcal{E}}(x) - \mathrm{Prox}_{\mathcal{E}}(y)\|^2 + \|(\mathrm{Id} - \mathrm{Prox}_{\mathcal{E}})(x) - (\mathrm{Id} - \mathrm{Prox}_{\mathcal{E}})(y)\|^2 \leqslant \|x - y\|^2$$

An Idea: Splitting

$$\min_x \mathcal{E}(x) = \underbrace{\frac{1}{2}\|y - \Phi x\|^2}_{F} + \underbrace{\lambda\|x\|_1}_{\lambda J}$$

$\mathcal{E}$ not smooth ☹ / $\text{Prox}_{\mathcal{E}}$ hard to compute ☹

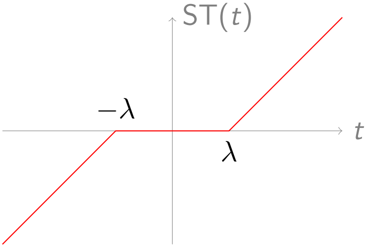
An Idea: Splitting

$$\min_x \mathcal{E}(x) = \underbrace{\frac{1}{2}\|y - \Phi x\|^2}_{F} + \underbrace{\lambda\|x\|_1}_{\lambda J}$$

$\mathcal{E}$ not smooth ☹ / $\text{Prox}_{\mathcal{E}}$ hard to compute ☹

But:

- F is smooth
- $\text{Prox}_{\lambda J}$ is easy to compute



Soft thresholding

$$(\text{Prox}_{\lambda\|\cdot\|_1}(x))_i = \text{sign}(x_i)(|x_i| - \lambda)_+$$

Fixed Point

$$x^\star \in \operatorname{argmin} f + g$$
$$0 \in \nabla F(x^\star) + \lambda \partial J(x^\star)$$

Fixed Point

$$
\begin{aligned}
x^\star &\in \operatorname{argmin} f + g \\
0 &\in \nabla F(x^\star) + \lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) + \rho \lambda \partial J(x^\star)
\end{aligned}
$$

Fixed Point

$$\begin{aligned} x^\star &\in \operatorname{argmin} f + g \\ 0 &\in \nabla F(x^\star) + \lambda \partial J(x^\star) \\ 0 &\in \rho \nabla F(x^\star) + \rho\lambda \partial J(x^\star) \\ 0 &\in \rho \nabla F(x^\star) - x^\star + x^\star + \rho\lambda \partial J(x^\star) \end{aligned}$$

Fixed Point

$$\begin{aligned} x^\star &\in \operatorname{argmin} f + g \\ 0 &\in \nabla F(x^\star) + \lambda \partial J(x^\star) \\ 0 &\in \rho \nabla F(x^\star) + \rho\lambda \partial J(x^\star) \\ 0 &\in \rho \nabla F(x^\star) - x^\star + x^\star + \rho\lambda \partial J(x^\star) \\ (\mathrm{Id} - \rho \nabla F)(x^\star) &\in (\mathrm{Id} + \rho\lambda \partial J)(x^\star) \end{aligned}$$

Fixed Point

$$
\begin{aligned}
x^\star &\in \operatorname{argmin} f + g \\
0 &\in \nabla F(x^\star) + \lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) + \rho\lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) - x^\star + x^\star + \rho\lambda \partial J(x^\star) \\
(\mathrm{Id} - \rho \nabla F)(x^\star) &\in (\mathrm{Id} + \rho\lambda \partial J)(x^\star) \\
x^\star &= (\mathrm{Id} + \rho\lambda \partial J)^{-1} (\mathrm{Id} - \rho \nabla F)(x^\star)
\end{aligned}
$$

Fixed Point

$$\begin{aligned}
x^\star &\in \operatorname{argmin} f + g \\
0 &\in \nabla F(x^\star) + \lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) + \rho\lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) - x^\star + x^\star + \rho\lambda \partial J(x^\star) \\
(\mathrm{Id} - \rho \nabla F)(x^\star) &\in (\mathrm{Id} + \rho\lambda \partial J)(x^\star) \\
x^\star &= (\mathrm{Id} + \rho\lambda \partial J)^{-1} (\mathrm{Id} - \rho \nabla F)(x^\star) \\
x^\star &= \mathrm{Prox}_{\rho\lambda J}(x^\star - \rho \nabla F(x^\star))
\end{aligned}$$

Fixed Point

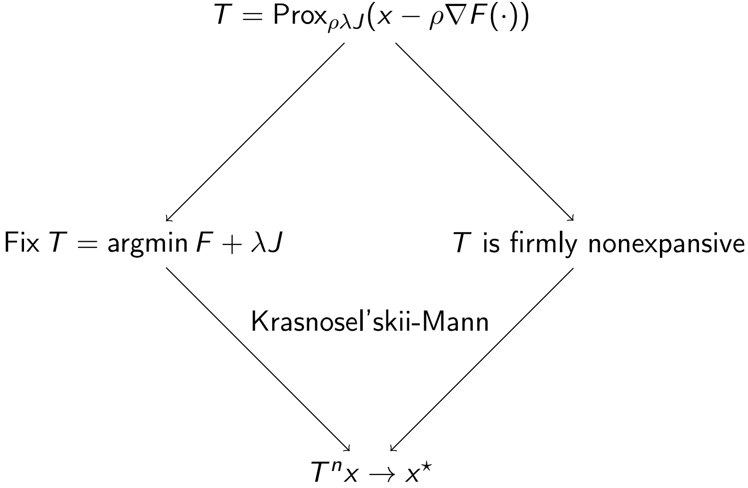
$$
\begin{aligned}
x^\star &\in \operatorname{argmin} f + g \\
0 &\in \nabla F(x^\star) + \lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) + \rho\lambda \partial J(x^\star) \\
0 &\in \rho \nabla F(x^\star) - x^\star + x^\star + \rho\lambda \partial J(x^\star) \\
(\mathrm{Id} - \rho \nabla F)(x^\star) &\in (\mathrm{Id} + \rho\lambda \partial J)(x^\star) \\
x^\star &= (\mathrm{Id} + \rho\lambda \partial J)^{-1} (\mathrm{Id} - \rho \nabla F)(x^\star) \\
x^\star &= \mathrm{Prox}_{\rho\lambda J}(x^\star - \rho \nabla F(x^\star))
\end{aligned}
$$

Proposition

$$Tx = \mathrm{Prox}_{\rho\lambda J}(x - \rho \nabla F(x))$$

$$\mathrm{Fix}\, T = \operatorname{argmin} F + \lambda J$$

Fixed Point


$T = \text{Prox}_{\rho\lambda J}(x - \rho\nabla F(\cdot))$
Fix $T = \text{argmin}\, F + \lambda J$
T is firmly nonexpansive
Krasnosel'skii-Mann
$T^n x \to x^\star$

Algorithm: Forward-Backward

$$x^{(n+1)} = \underbrace{\mathrm{Prox}_{\rho\lambda J}}_{\text{backward}}(\underbrace{x^{(n)} - \rho\nabla F(x^{(n)})}_{\text{forward}})$$

Special cases

- Gradient descent: $J = 0$

$$x^{(n+1)} = x^{(n)} - \rho\nabla F(x^{(n)})$$

Algorithm: Forward-Backward

$$x^{(n+1)} = \underbrace{\mathrm{Prox}_{\rho\lambda J}}_{\text{backward}}(\underbrace{x^{(n)} - \rho\nabla F(x^{(n)})}_{\text{forward}})$$

Special cases

- Gradient descent: $J = 0$

$$x^{(n+1)} = x^{(n)} - \rho\nabla F(x^{(n)})$$

- Proximal point: $F = 0$

$$x^{(n+1)} = \mathrm{Prox}_{\rho\lambda J}(x^{(n)})$$

Algorithm: Forward-Backward

$$x^{(n+1)} = \underbrace{\mathrm{Prox}_{\rho\lambda J}}_{\text{backward}}(\underbrace{x^{(n)} - \rho\nabla F(x^{(n)})}_{\text{forward}})$$

Special cases

- Gradient descent: $J = 0$

$$x^{(n+1)} = x^{(n)} - \rho\nabla F(x^{(n)})$$

- Proximal point: $F = 0$

$$x^{(n+1)} = \mathrm{Prox}_{\rho\lambda J}(x^{(n)})$$

- Projected gradient: $J = \iota_{\mathcal{C}}$

$$x^{(n+1)} = \Pi_{\mathcal{C}}(x^{(n)} - \rho\nabla F(x^{(n)}))$$

Model Identifiability

$$x^{(k+1)} = \mathrm{Prox}_{\rho\lambda J}(x^{(k)} - \rho \nabla F(x^{(k)}))$$

Theorem

Assume that $J \in \mathrm{PS}_{x_0}(\mathcal{M})$, $(\mathrm{INJ}_{T_{x_0}})$, $\Phi^* \eta_F \in \mathrm{ri}\, \partial J(x_0)$ *hold and* $0 < \rho < 2/\|\Phi\|$. *Then, there exist* (C, C') *such that if* $\|w\|_2 \leqslant C$ *and* $\lambda = C'\|w\|_2$, *there exists* k_0 *such that for all* $k \geqslant k_0$

$$x^{(k)} \in \mathcal{M} \quad \text{and} \quad x^{(k)} \to x_0.$$

Model Identifiability

$$x^{(k+1)} = \mathrm{Prox}_{\rho\lambda J}(x^{(k)} - \rho\nabla F(x^{(k)}))$$

Theorem

Assume that $J \in \mathrm{PS}_{x_0}(\mathcal{M})$, $(\mathrm{INJ}_{T_{x_0}})$, $\Phi^*\eta_F \in \mathrm{ri}\,\partial J(x_0)$ *hold and* $0 < \rho < 2/\|\Phi\|$. *Then, there exist* (C, C') *such that if* $\|w\|_2 \leqslant C$ *and* $\lambda = C'\|w\|_2$, *there exists* k_0 *such that for all* $k \geqslant k_0$

$$x^{(k)} \in \mathcal{M} \quad \textit{and} \quad x^{(k)} \to x_0.$$

Not a rate of convergence result but a finite time result

Take Away

- Unified analysis of recovery guarantees for regularized linear inverse problems
- Partial smoothness provides a nice framework to work with

Take Away

- Unified analysis of recovery guarantees for regularized linear inverse problems
- Partial smoothness provides a nice framework to work with

Several (semi-) open problems:

- How to define a functional J from a set of model $(\mathcal{M})_{\mathcal{M}\in\mathbb{M}}$?
- What happens for J not convex?
- What occurs at the boundary ?
- In infinite dimension ?

Take Away

- Unified analysis of recovery guarantees for regularized linear inverse problems
- Partial smoothness provides a nice framework to work with

Several (semi-) open problems:

- How to define a functional J from a set of model $(\mathcal{M})_{\mathcal{M} \in \mathbb{M}}$?
- What happens for J not convex?
- What occurs at the boundary ?
- In infinite dimension ?

Thanks for your attention!

More?

- S. V., G. Peyré, and J. Fadili
  Low Complexity Regularization of Linear Inverse Problems
  Sampling Theory, a Renaissance, 2015
  $\rightarrow$ review chapter (cover the same spectrum of topics)
- S. V., C. Deledalle, G. Peyré, J. Fadili, and C. Dossal
  The Degrees of Freedom of Partly Smooth Regularizers
  Annals of the Institute of Statistical Mathematics, 2016
  $\rightarrow$ risk estimation and sensitivity
- S. V., G. Peyré, and J. Fadili
  Model Consistency of Partly Smooth Regularizers
  preprint (HAL/arxiv), 2014
  $\rightarrow$ model stability & identifiability
- J. Fadili, G. Peyré, S. Vaiter, C. Deledalle, and J. Salmon
  Stable Recovery with Analysis Decomposable Priors
  Proc. SampTA, 2013
  $\rightarrow$ ℓ^2-stability

Part V: Parameter Selection

In order to solve

$$y = \Phi x_0 + w$$

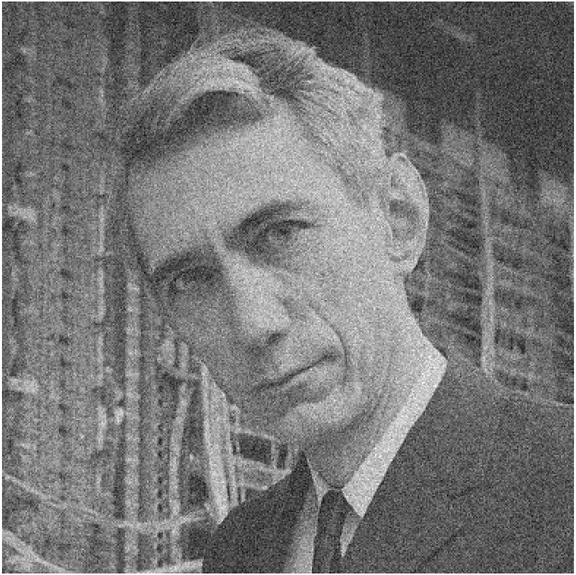
we consider for a partly smooth function J, the optimization

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

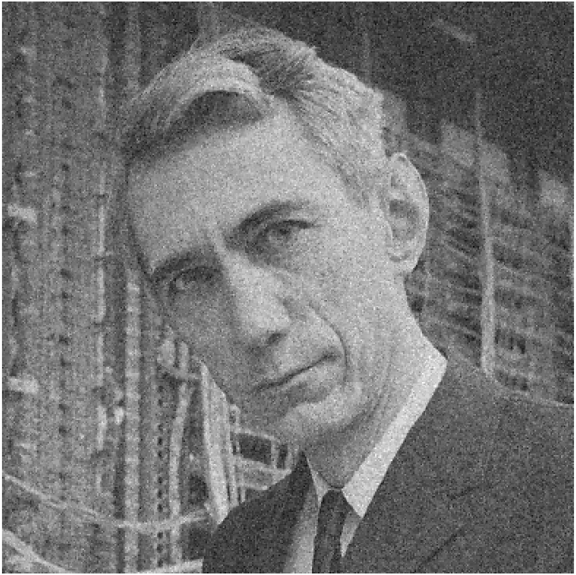
Goal

How to choose an adequate λ ?

Influence of λ

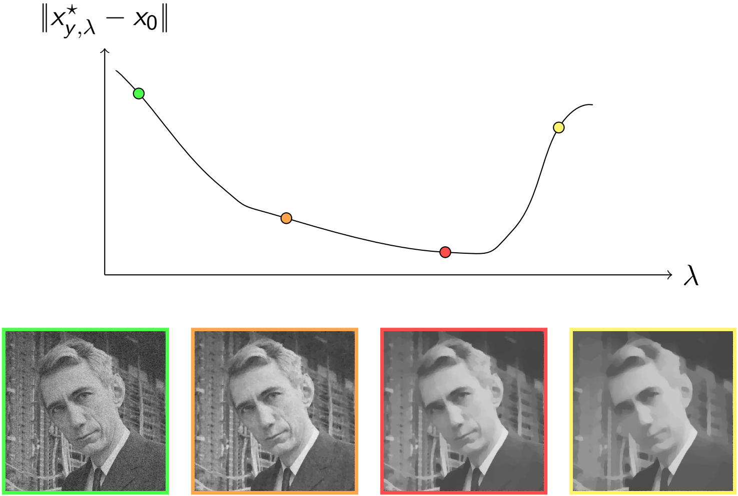


Influence of λ



Influence of λ

Influence of λ



Let's be Random

Until now, we considered deterministic observations

$$y = \Phi x_0 + w$$

Let's add a noise model, for instance a Gaussian one

$$Y = \Phi x_0 + W \quad \text{where} \quad W \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_q)$$

Risk Estimation

$$x^\star_{y,\lambda} \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}}\ \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Risk Estimation

$$x^\star_\lambda(y) \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \ \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Prediction risk

$$R_\lambda(Y) = \mathbb{E}_W \|\Phi x^\star_\lambda(Y) - \Phi x_0\|_2^2$$

Our objective is to minimize this risk, i.e. finding

$$\lambda^\star(Y) \in \underset{\lambda \in \mathbb{R}^*_+}{\text{Argmin}} \ R_\lambda(Y)$$

Issue

In practice, x_0 is not known ...

We are going to define an estimator of $R_\lambda(Y)$

Degrees of Freedom and Stein's Lemma

Simple Example

$$x^\star_\lambda(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

If J is smooth, first-order condition:

$$\Phi^*(\Phi x^\star_\lambda(y) - y) + \lambda \nabla J(x^\star_\lambda(y)) = 0$$

If $\Gamma = \Phi^*\Phi + \lambda \mathrm{D}^2 J(x^\star_\lambda(y))$ is invertible, implicit function theorem gives

$$\mathrm{D}x^\star_\lambda(y) = \Gamma^{-1}\Phi^*$$