Low Complexity Regularizations of Inverse Problems

Samuel Vaiter
CEREMADE, Univ. Paris-Dauphine
vaiter@ceremade.dauphine.fr

July 10, 2014

PhD defense

Linear Inverse Problems

Forward model

$$y = \Phi\, x_0 + w$$

Forward operator $\Phi : \mathbb{R}^n \to \mathbb{R}^q$ linear ($q \leqslant n$)

Linear Inverse Problems

Forward model

$$y = \Phi x_0 + w$$

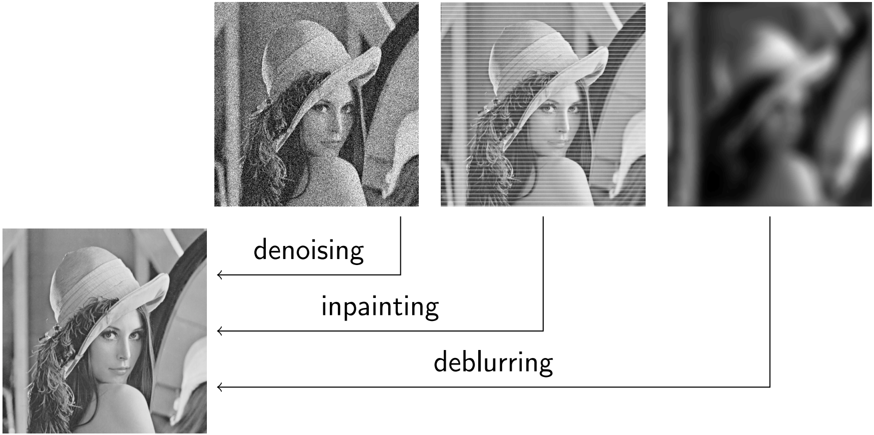
Forward operator $\Phi : \mathbb{R}^n \to \mathbb{R}^q$ linear ($q \leqslant n$) $\to$ ill-posed problem

Linear Inverse Problems

Forward model

$$y = \Phi\, x_0 + w$$

Forward operator $\Phi : \mathbb{R}^n \to \mathbb{R}^q$ linear ($q \leqslant n$) $\to$ ill-posed problem



Variational Regularization

Trade-off between prior regularization and data fidelity

Variational Regularization

Trade-off between prior regularization and data fidelity

$$x^\star \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \; J(x) + \frac{1}{2\lambda} \|y - \Phi x\|^2 \qquad (P_{y,\lambda})$$

Variational Regularization

Trade-off between prior regularization and data fidelity

$$x^\star \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}}\; J(x) + \frac{1}{2\lambda} \|y - \Phi x\|^2 \qquad (P_{y,\lambda})$$

$$\Big\downarrow \lambda \to 0^+$$

$$x^\star \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}}\; J(x) \quad \text{subject to} \quad y = \Phi x \qquad (P_{y,0})$$

Variational Regularization

Trade-off between prior regularization and data fidelity

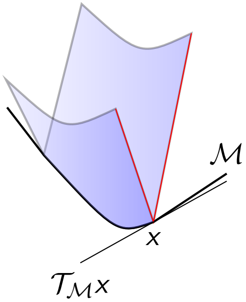
$$x^\star \in \underset{x \in \mathbb{R}^{\boldsymbol{n}}}{\mathrm{Argmin}} \; J(x) + \frac{1}{2\lambda} \|y - \Phi x\|^2 \qquad (P_{y,\lambda})$$

$$\Big\downarrow \lambda \to 0^+$$

$$x^\star \in \underset{x \in \mathbb{R}^{\boldsymbol{n}}}{\mathrm{Argmin}} \; J(x) \quad \text{subject to} \quad y = \Phi x \qquad (P_{y,0})$$

J convex, bounded from below and finite-valued function, typically non-smooth.

Partly Smooth Functions [Lewis 2002]



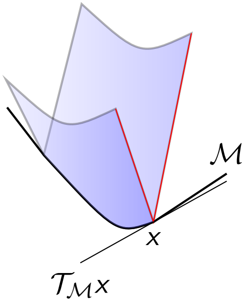
J is *partly smooth* at x relative to a C^2-manifold $\mathcal{M}$ if

Smoothness. J restricted to $\mathcal{M}$ is C^2 around x

Sharpness. $\forall h \in (\mathcal{T}_{\mathcal{M}}x)^{\perp}$, $t \mapsto J(x+th)$ is non-smooth at $t = 0$.

Continuity. ∂J on $\mathcal{M}$ is continuous around x.

Partly Smooth Functions [Lewis 2002]



J is *partly smooth* at x relative to a C^2-manifold $\mathcal{M}$ if

Smoothness. J restricted to $\mathcal{M}$ is C^2 around x

Sharpness. $\forall h \in (\mathcal{T}_{\mathcal{M}} x)^{\perp}$, $t \mapsto J(x+th)$ is non-smooth at $t=0$.

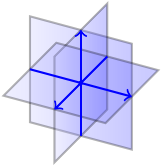
Continuity. ∂J on $\mathcal{M}$ is continuous around x.

$$J, G \text{ partly smooth} \Rightarrow \begin{cases} J+G \\ J \circ D^* \text{ with } D \text{ linear operator} \\ J \circ \sigma \text{ (spectral lift)} \end{cases} \quad \text{partly smooth}$$

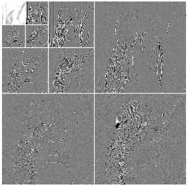
Low Complexity Models

Sparsity

$$J(x) = \sum_{i=1,\ldots,n} |x_i|$$



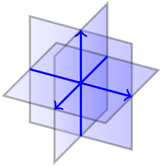
$$\mathcal{M}_x = \left\{ x' : \ \mathrm{supp}(x') \subseteq \mathrm{supp}(x) \right\}$$



Low Complexity Models

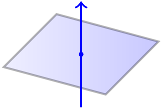
Sparsity

$$J(x) = \sum_{i=1,\ldots,n} |x_i|$$

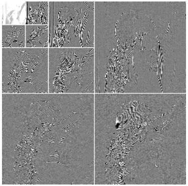


Group sparsity

$$J(x) = \sum_{b \in \mathcal{B}} \|x_b\|$$



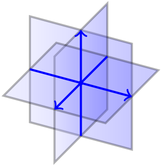
$$\mathcal{M}_x = \left\{ x' : \operatorname{supp}(x') \subseteq \operatorname{supp}(x) \right\}$$



Low Complexity Models

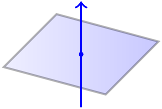
Sparsity

$$J(x) = \sum_{i=1,\ldots,n} |x_i|$$



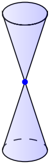
Group sparsity

$$J(x) = \sum_{b \in \mathcal{B}} \|x_b\|$$

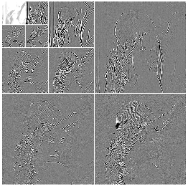


Low rank

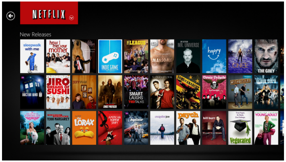
$$J(x) = \sum_{i=1,\ldots,n} |\sigma_i(x)|$$



$$\mathcal{M}_x = \{x' \,:\, \mathrm{supp}(x') \subseteq \mathrm{supp}(x)\}$$



$$\mathcal{M}_x = \{x' \,:\, \mathrm{rank}(x') = \mathrm{rank}(x)\}$$



Contributions of the Thesis

$$x^\star \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

I. Noise ℓ^2 Stability

Ensure that

$$\|x^\star - x_0\|_2 = O(\|w\|_2)$$

II. Stable Manifold Identification

Ensure that

$$x^\star \in \mathcal{M}_{x_0}$$

III. Local Differentiability

Variation of $x^\star$ as a function of the observations y.

IV. Unbiased Risk Estimation

Select the optimal parameter λ.

Contributions of the Thesis

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

I. Noise ℓ^2 Stability

Ensure that

$$\|x^\star - x_0\|_2 = O(\|w\|_2)$$

II. Stable Manifold Identification

Ensure that

$$x^\star \in \mathcal{M}_{x_0}$$

III. Local Differentiability

Variation of $x^\star$ as a function of the observations y.

IV. Unbiased Risk Estimation

Select the optimal parameter λ.

Variations on a theme:
Sensitivity Analysis of Optimization Problems

Contributions of the Thesis

$$x^\star \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

I. Noise ℓ^2 Stability

Ensure that

$$\|x^\star - x_0\|_2 = O(\|w\|_2)$$

II. Stable Manifold Identification

Ensure that

$$x^\star \in \mathcal{M}_{x_0}$$

III. Local Differentiability

Variation of $x^\star$ as a function of the observations y.

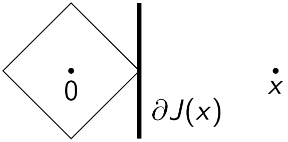
IV. Unbiased Risk Estimation

Select the optimal parameter λ.

Variations on a theme:
Sensitivity Analysis of Optimization Problems

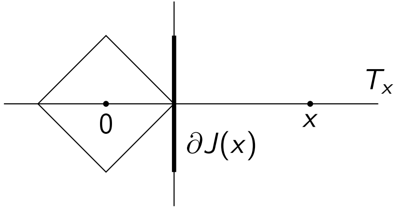
Model Tangent Subspace

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^n : \; \forall x', J(x) - J(x') \geqslant \langle \eta, \, x - x' \rangle \right\}$$



Model Tangent Subspace

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^n : \ \forall x', J(x) - J(x') \geqslant \langle \eta, \, x - x' \rangle \right\}$$

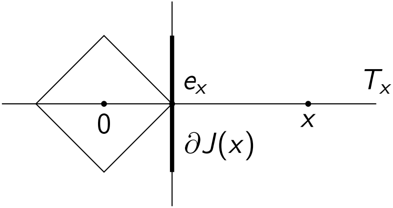


Model tangent space $T_x = \mathrm{VectHull}(\partial J(x))^\perp$

If J partly smooth at x for $\mathcal{M}$ and $\mathcal{M}$ is linear, then $\mathcal{T}_{\mathcal{M}} x = T_x$

Model Tangent Subspace

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^n : \ \forall x', J(x) - J(x') \geqslant \langle \eta, \, x - x' \rangle \right\}$$



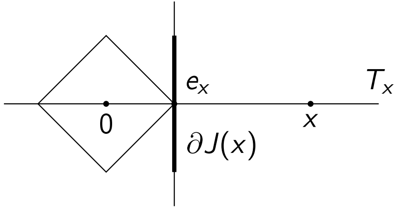
Model tangent space $T_x = \mathrm{VectHull}(\partial J(x))^\perp$

Model vector $e_x = \mathrm{P}_{T_x}(\partial J(x))$

If J partly smooth at x for $\mathcal{M}$ and $\mathcal{M}$ is linear, then $\mathcal{T}_{\mathcal{M}} x = T_x$

Model Tangent Subspace

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^n : \ \forall x', J(x) - J(x') \geqslant \langle \eta,\, x - x' \rangle \right\}$$



Model tangent space $T_x = \mathrm{VectHull}(\partial J(x))^\perp$

Model vector $e_x = \mathrm{P}_{T_x}(\partial J(x))$

If J partly smooth at x for $\mathcal{M}$ and $\mathcal{M}$ is linear, then $\mathcal{T}_{\mathcal{M}} x = T_x$

Sparsity $\|\cdot\|_1$

$$T_x = \{\eta : \ \mathrm{supp}(\eta) \subseteq \mathrm{supp}(x)\}$$
$$e_x = \mathrm{sign}(x)$$

Trace Norm $\|\cdot\|_*$

$$T_x = \{\eta : \ U_\perp^* \eta V_\perp = 0\}$$
$$e_x = UV^*$$

Dual Certificates

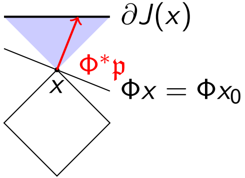
$$x^\star \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}}\ J(x) \quad \text{subject to} \quad y = \Phi x \qquad (P_{y,0})$$

Dual Certificates

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}}\ J(x) \quad \text{subject to} \quad y = \Phi x \qquad\qquad (P_{y,0})$$

Source condition

$$\Phi^* \mathfrak{p} \in \partial J(x)$$

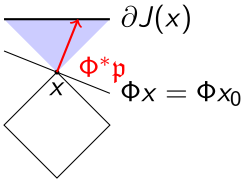


Dual Certificates

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}}\ J(x) \quad \text{subject to} \quad y = \Phi x \qquad (P_{y,0})$$

Source condition

$$\Phi^* \mathfrak{p} \in \partial J(x)$$



Proposition

There exists a dual certificate $\mathfrak{p}$ if, and only if, x_0 is a solution of $(\mathcal{P}_{y,0})$.

Dual Certificates

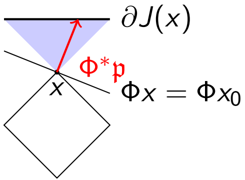
$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}}\ J(x) \quad \text{subject to} \quad y = \Phi x \qquad (P_{y,0})$$

Source condition

$$\Phi^* \mathfrak{p} \in \partial J(x)$$

Non-degenerate source condition

$$\Phi^* \mathfrak{p} \in \operatorname{ri} \partial J(x)$$



Proposition

There exists a dual certificate $\mathfrak{p}$ if, and only if, x_0 is a solution of $(\mathcal{P}_{y,0})$.

Linear Convergence Rate

$$y = \Phi x_0 + w$$

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \ \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

Theorem

Assume

$$\Phi^* \mathfrak{p} \in \operatorname{ri} \partial J(x_0) \quad \text{and} \quad \operatorname{Ker} \Phi \cap T_{x_0} = \{0\}.$$

Choosing $\lambda = c\|w\|_2$, $c > 0$, for any minimizer $x^\star$ of $(P_{y,\lambda})$

$$\|x^\star - x_0\|_2 \leqslant C(c, \mathfrak{p})\|w\|_2 \ .$$

Linear Convergence Rate

$$y = \Phi x_0 + w$$

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \ \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

Theorem

Assume

$$\Phi^* \mathfrak{p} \in \operatorname{ri} \partial J(x_0) \quad \text{and} \quad \operatorname{Ker} \Phi \cap T_{x_0} = \{0\}.$$

Choosing $\lambda = c\|w\|_2$, $c > 0$, for any minimizer $x^\star$ of $(P_{y,\lambda})$

$$\|x^\star - x_0\|_2 \leqslant C(c, \mathfrak{p})\|w\|_2 \ .$$

[Grasmair et al. 2010]: ℓ^1

[Grasmair 2011]: $J(x^\star - x_0) = O(\|w\|_2)$

Contributions of the Thesis

$$x^\star \in \underset{x\in\mathbb{R}^n}{\text{Argmin}}\ \frac{1}{2}\|y-\Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

I. Noise ℓ^2 Stability

Ensure that

$$\|x^\star - x_0\|_2 = O(\|w\|_2)$$

II. Stable Manifold Identification

Ensure that

$$x^\star \in \mathcal{M}_{x_0}$$

III. Local Differentiability

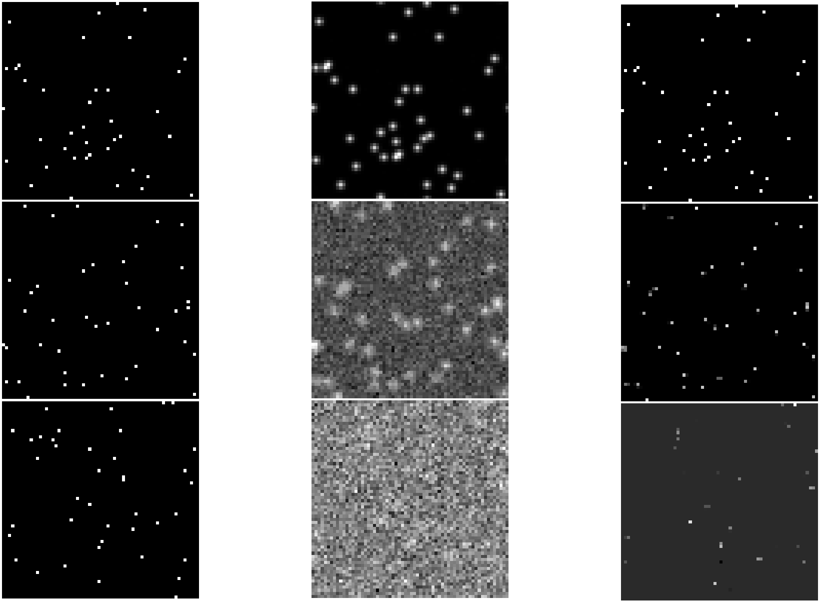
Variation of $x^\star$ as a function of the observations y.

IV. Unbiased Risk Estimation

Select the optimal parameter λ.

Variations on a theme:
Sensitivity Analysis of Optimization Problems

Objective



x_0 y $x^\star$

Linearized Precertificate

Minimal norm certificate

$$\mathfrak{p}_0 = \operatorname{argmin} \|\mathfrak{p}\| \quad \text{subject to} \quad \Phi^*\mathfrak{p} \in \partial J(x_0)$$

Linearized Precertificate

Minimal norm certificate

$$\mathfrak{p}_0 = \operatorname{argmin} \|\mathfrak{p}\| \quad \text{subject to} \quad \Phi^* \mathfrak{p} \in \partial J(x_0)$$

Linearized precertificate

$$\mathfrak{p}_F = \operatorname{argmin} \|\mathfrak{p}\| \quad \text{subject to} \quad \Phi^* \mathfrak{p} \in \operatorname{aff} \partial J(x_0) \Leftrightarrow (\Phi^* \mathfrak{p})_{T_{x_0}} = e_{x_0}$$

Linearized Precertificate

Minimal norm certificate

$$\mathfrak{p}_0 = \operatorname{argmin} \|\mathfrak{p}\| \quad \text{subject to} \quad \Phi^* \mathfrak{p} \in \partial J(x_0)$$

Linearized precertificate

$$\mathfrak{p}_F = \operatorname{argmin} \|\mathfrak{p}\| \quad \text{subject to} \quad \Phi^* \mathfrak{p} \in \operatorname{aff} \partial J(x_0) \Leftrightarrow (\Phi^* \mathfrak{p})_{T_{x_0}} = e_{x_0}$$

Proposition

Assume $\operatorname{Ker} \Phi \cap T = \{0\}$ where $T = T_{x_0}$. Then,

$$\mathfrak{p}_F = \Phi_T^{+,*} e$$

where $\Phi_T = \Phi P_T$. Moreover,

$$\mathfrak{p}_F \in \operatorname{ri} \partial J(x_0) \Rightarrow \mathfrak{p}_F = \mathfrak{p}_0$$

Manifold Selection

Theorem

Assume J is partly smooth at x_0 relative to $\mathcal{M}$. If

$$\Phi^* \mathfrak{p}_F \in \operatorname{ri} \partial J(x_0) \quad \text{and} \quad \operatorname{Ker} \Phi \cap T_{x_0} = \{0\}.$$

There exists $C > 0$ such that if

$$\max(\lambda, \|w\|/\lambda) \leqslant C,$$

the unique solution $x^\star$ of $(P_{y,\lambda})$ satisfies

$$x^\star \in \mathcal{M} \quad \text{and} \quad \|x^\star - x_0\| = O(\|w\|).$$

Manifold Selection

Theorem

Assume J is partly smooth at x_0 relative to $\mathcal{M}$. If

$$\Phi^* \mathfrak{p}_F \in \operatorname{ri} \partial J(x_0) \quad \text{and} \quad \operatorname{Ker} \Phi \cap T_{x_0} = \{0\}.$$

There exists $C > 0$ such that if

$$\max(\lambda, \|w\|/\lambda) \leqslant C,$$

the unique solution $x^\star$ of $(P_{y,\lambda})$ satisfies

$$x^\star \in \mathcal{M} \quad \text{and} \quad \|x^\star - x_0\| = O(\|w\|).$$

Almost sharp analysis ($\Phi^* \mathfrak{p}_F \notin \partial J(x_0) \Rightarrow x^\star \notin \mathcal{M}_{x_0}$)

Manifold Selection

Theorem

Assume J is partly smooth at x_0 relative to $\mathcal{M}$. If

$$\Phi^* \mathfrak{p}_F \in \mathrm{ri}\,\partial J(x_0) \quad \text{and} \quad \mathrm{Ker}\,\Phi \cap T_{x_0} = \{0\}.$$

There exists $C > 0$ such that if

$$\max(\lambda, \|w\|/\lambda) \leqslant C,$$

the unique solution $x^\star$ of $(P_{y,\lambda})$ satisfies

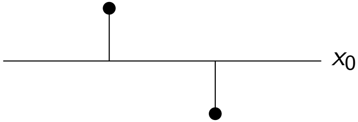
$$x^\star \in \mathcal{M} \quad \text{and} \quad \|x^\star - x_0\| = O(\|w\|).$$

Almost sharp analysis ($\Phi^* \mathfrak{p}_F \notin \partial J(x_0) \Rightarrow x^\star \notin \mathcal{M}_{x_0}$)

[Fuchs 2004]: ℓ^1

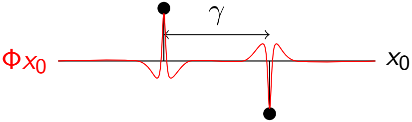
[Bach 2008]: $\ell^1 - \ell^2$ and nuclear norm.

Sparse Spike Deconvolution



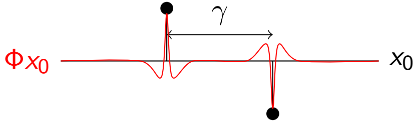
Sparse Spike Deconvolution

$$\Phi x = \sum_i x_i \varphi(\cdot - \Delta i) \quad J(x) = \|x\|_1$$



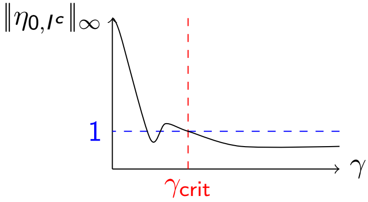
Sparse Spike Deconvolution

$$\Phi x = \sum_i x_i \varphi(\cdot - \Delta i) \quad J(x) = \|x\|_1$$



$$\Phi^* \eta_F \in \operatorname{ri} \partial J(x_0) \Leftrightarrow \|\Phi_{I^c}^{+,*} \Phi_I \operatorname{sign}(x_{0,I})\|_\infty < 1 \Leftrightarrow \text{stable recovery}$$

$$I = \operatorname{supp}(x_0)$$

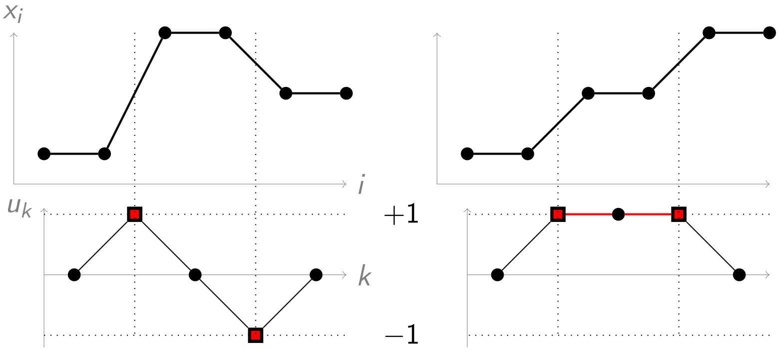


1D Total Variation and Jump Set

$$J = \|\nabla_d \cdot \|_1, \mathcal{M}_x = \left\{x' : \ \mathsf{supp}(\nabla_d x') \subseteq \mathsf{supp}(\nabla_d x)\right\}, \Phi = \mathrm{Id}$$

1D Total Variation and Jump Set

$$J = \|\nabla_d \cdot\|_1, \mathcal{M}_x = \left\{x' : \operatorname{supp}(\nabla_d x') \subseteq \operatorname{supp}(\nabla_d x)\right\}, \Phi = \mathrm{Id}$$



stable jump

unstable jump

$$\Phi^* \mathfrak{p}_F = \operatorname{div} u$$

Contributions of the Thesis

$$x^\star \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \ \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

I. Noise ℓ^2 Stability

Ensure that

$$\|x^\star - x_0\|_2 = O(\|w\|_2)$$

II. Stable Manifold Identification

Ensure that

$$x^\star \in \mathcal{M}_{x_0}$$

III. Local Differentiability

Variation of $x^\star$ as a function of the observations y.

IV. Unbiased Risk Estimation

Select the optimal parameter λ.

Variations on a theme:
Sensitivity Analysis of Optimization Problems

Objective

$$x^\star \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \ \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

If J is smooth, first-order conditions

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

Objective

$$x^\star \in \underset{x\in\mathbb{R}^n}{\text{Argmin}}\ \frac{1}{2}\|y-\Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

If J is smooth, first-order conditions

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

Implicit function theorem $\to$ If

$$\Gamma = \Phi^*\Phi + \lambda \mathrm{D}^2 J(x^\star)$$

is invertible, then on a neighborhood, there exists $\bar{y} \mapsto \tilde{x}(\bar{y})$ (C^1) such that $\tilde{x}(\bar{y})$ solution of $(P_{\bar{y},\lambda})$, $\tilde{x}(y) = x^\star$ and

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

Objective

$$x^\star \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}} \ \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

If J is smooth, first-order conditions

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

Implicit function theorem $\rightarrow$ If

$$\Gamma = \Phi^*\Phi + \lambda \mathrm{D}^2 J(x^\star)$$

is invertible, then on a neighborhood, there exists $\bar{y} \mapsto \tilde{x}(\bar{y})$ (C^1) such that $\tilde{x}(\bar{y})$ solution of $(P_{\bar{y},\lambda})$, $\tilde{x}(y) = x^\star$ and

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

Issues

1. What if J not differentiable ?
2. What if Γ is not invertible ?

Transition Space

J partly smooth with linear manifolds and $(T_x)_{x \in \mathbb{R}^n}$ finite

$$\mathcal{H} = \bigcup_{T \in (T_x)_{x \in \mathbb{R}^n}} \mathrm{bd}\left(\Pi_1 \left\{ (y, x_T) : \frac{1}{\lambda} \Phi_T^* (\Phi_T x_T - y) \in \mathrm{rbd}\, \partial J(x_T) \right\} \right)$$

Π_1: projection on the first component

Transition Space

J partly smooth with linear manifolds and $(T_x)_{x\in\mathbb{R}^n}$ finite

$$\mathcal{H} = \bigcup_{T\in(T_x)_{x\in\mathbb{R}^n}} \mathrm{bd}\left(\Pi_1\left\{(y, x_T) :\ \frac{1}{\lambda}\Phi_T^*(\Phi_T x_T - y) \in \mathrm{rbd}\,\partial J(x_T)\right\}\right)$$

Π_1: projection on the first component

Lemma

If J is semi-algebraic, $\mathcal{H}$ is semi-algebraic and has zero measure w.r.t Lebesgue measure on $\mathbb{R}^q$.

Semi-algebraic set: boolean combination of polynomial equalities and inequalities
Semi-algebraic function: function with a semi-algebraic graph

Transition Space

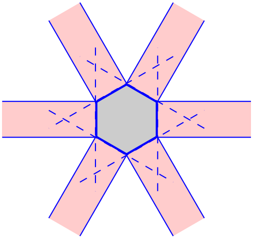
J partly smooth with linear manifolds and $(T_x)_{x\in\mathbb{R}^n}$ finite

$$\mathcal{H} = \bigcup_{T\in(T_x)_{x\in\mathbb{R}^n}} \mathrm{bd}\left(\Pi_1\left\{(y, x_T) :\ \frac{1}{\lambda}\Phi_T^*(\Phi_T x_T - y) \in \mathrm{rbd}\,\partial J(x_T)\right\}\right)$$

Π_1: projection on the first component

Lemma

If J is semi-algebraic, $\mathcal{H}$ is semi-algebraic and has zero measure w.r.t Lebesgue measure on $\mathbb{R}^q$.



$J = \|\cdot\|_1, \Phi \in \mathbb{R}^{2\times 3}$

Semi-algebraic set: boolean combination of polynomial equalities and inequalities

Semi-algebraic function: function with a semi-algebraic graph

ex: polyhedral functions $\Rightarrow$ $\mathcal{H}$ is a union of subspaces

Local Behavior

$$x^\star \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

Theorem

Let $T = T_{x^\star}$. Assume

$$y \notin \mathcal{H} \quad \text{and} \quad \operatorname{Ker}\Phi \cap \operatorname{Ker} \mathrm{D}^2 J_T(x^\star) = \{0\}.$$

Then, there exists, on a neighborhood $\mathcal{O}$ of y, a mapping $\tilde{x}$ s.t.

1. For every $\bar{y} \in \mathcal{O}$, $\tilde{x}(\bar{y})$ is a solution of $(\mathcal{P}_{\bar{y},\lambda})$ and $\tilde{x}(y) = x^\star$.
2. The mapping $\tilde{x}$ is $\mathrm{C}^1(\mathcal{O})$ and

$$\mathrm{D}_1 \tilde{x}(y) = (\Phi_T^* \Phi_T + \lambda \mathrm{D}^2 J_T(x^\star))^{-1} \Phi_T$$

Local Behavior

$$x^\star \in \underset{x\in\mathbb{R}^n}{\operatorname{Argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

Theorem

Let $T = T_{x^\star}$. Assume

$$y \notin \mathcal{H} \quad \text{and} \quad \operatorname{Ker}\Phi \cap \operatorname{Ker} \mathrm{D}^2 J_T(x^\star) = \{0\}.$$

Then, there exists, on a neighborhood $\mathcal{O}$ of y, a mapping $\tilde{x}$ s.t.

1. For every $\bar{y} \in \mathcal{O}$, $\tilde{x}(\bar{y})$ is a solution of $(\mathcal{P}_{\bar{y},\lambda})$ and $\tilde{x}(y) = x^\star$.
2. The mapping $\tilde{x}$ is $C^1(\mathcal{O})$ and

$$\mathrm{D}_1\tilde{x}(y) = (\Phi_T^*\Phi_T + \lambda \mathrm{D}^2 J_T(x^\star))^{-1}\Phi_T$$

Lemma

There exists a solution $x^\star$ of $(P_{y,\lambda})$ s.t. $\operatorname{Ker}\Phi \cap \operatorname{Ker}\mathrm{D}^2 J_T(x^\star) = \{0\}$

Contributions of the Thesis

$$x^\star \in \underset{x \in \mathbb{R}^n}{\mathrm{Argmin}} \; \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (P_{y,\lambda})$$

I. Noise ℓ^2 Stability

Ensure that

$$\|x^\star - x_0\|_2 = O(\|w\|_2)$$

II. Stable Manifold Identification

Ensure that

$$x^\star \in \mathcal{M}_{x_0}$$

III. Local Differentiability

Variation of $x^\star$ as a function of the observations y.

IV. Unbiased Risk Estimation

Select the optimal parameter λ.

Variations on a theme:
Sensitivity Analysis of Optimization Problems

Objective

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

Stochastic linear model

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

Objective

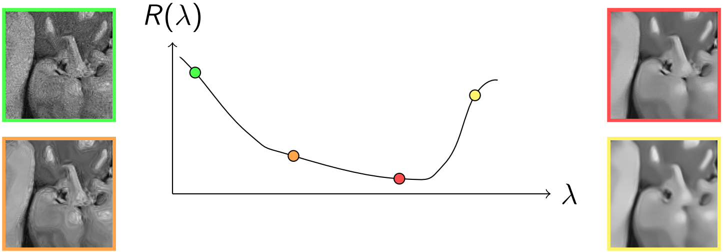
$$x^\star(y) \in \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \; \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \qquad (\mathcal{P}_{y,\lambda})$$

Stochastic linear model

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

(Estimation) risk

$$R(\lambda) = \mathbb{E}_W \left[\|x^\star(Y) - x_0\|_2^2 \right]$$



Stein's Unbiased Risk Estimation

Prediction : $\mu(y) = \Phi x^\star(y)$ (single-valued even if $x^\star$ not)

Stein's Unbiased Risk Estimation

Prediction : $\mu(y) = \Phi x^\star(y)$ (single-valued even if $x^\star$ not)

(Empirial) Degrees of freedom [Efròn 1986] : $df(Y) = \mathrm{div}(\mu)(Y)$

Stein's Unbiased Risk Estimation

Prediction : $\mu(y) = \Phi x^\star(y)$ (single-valued even if $x^\star$ not)

(Empirial) Degrees of freedom [Efròn 1986] : $df(Y) = \mathrm{div}(\mu)(Y)$

Stein's Unbiased Risk Estimator

$$\mathrm{SURE}(\mu)(Y) = \| Y - \mu(Y) \|^2 + 2\sigma^2 df(Y) - n\sigma^2$$

$\longrightarrow$ does not depend on the knowledge of x_0

Stein's Unbiased Risk Estimation

Prediction : $\mu(y) = \Phi x^\star(y)$ (single-valued even if $x^\star$ not)

(Empirial) Degrees of freedom [Efròn 1986] : $df(Y) = \mathrm{div}(\mu)(Y)$

Stein's Unbiased Risk Estimator

$$\mathrm{SURE}(\mu)(Y) = \| Y - \mu(Y) \|^2 + 2\sigma^2 df(Y) - n\sigma^2$$

$\longrightarrow$ does not depend on the knowledge of x_0

Lemma ([Stein 1981], restricted version)

If μ is Lipschitz, then

$$\mathbb{E}_W[\mathrm{SURE}(\mu)(Y)] = \mathbb{E}_W[\|\mu(Y) - \Phi x_0\|^2].$$

Stein's Unbiased Risk Estimation

Prediction : $\mu(y) = \Phi x^\star(y)$ (single-valued even if $x^\star$ not)

(Empirial) Degrees of freedom [Efròn 1986] : $df(Y) = \mathrm{div}(\mu)(Y)$

Stein's Unbiased Risk Estimator

$$\mathrm{SURE}(\mu)(Y) = \| Y - \mu(Y) \|^2 + 2\sigma^2 df(Y) - n\sigma^2$$

$\longrightarrow$ does not depend on the knowledge of x_0

Lemma ([Stein 1981], restricted version)

If μ is Lipschitz, then

$$\mathbb{E}_W[\mathrm{SURE}(\mu)(Y)] = \mathbb{E}_W[\|\mu(Y) - \Phi x_0\|^2].$$

$\rightarrow$ must have $\mathrm{div}(\mu)$ in closed form Lebesgue-almost everywhere

Practical Computation

Theorem

If J is semi-algebraic, then

$$df(Y) = \operatorname{tr}\left[\Phi_T(\Phi_T^*\Phi_T + \lambda \mathrm{D}^2 J_T(x^\star(Y)))^{-1}\Phi_T\right] \quad \text{a.e.}$$

where $T = T_{x^\star}$ and $x^\star$ a solution s.t. $\operatorname{Ker}\Phi \cap \operatorname{Ker}\mathrm{D}^2 J_T(x^\star) = \{0\}$.

$J = \|\cdot\|_1 \to df(Y) = |\operatorname{supp}(x^\star(Y))|$

Practical Computation

Theorem

If J is semi-algebraic, then

$$df(Y) = \operatorname{tr}\left[\Phi_T(\Phi_T^*\Phi_T + \lambda \mathrm{D}^2 J_T(x^\star(Y)))^{-1}\Phi_T\right] \quad \text{a.e.}$$

where $T = T_{x^\star}$ and $x^\star$ a solution s.t. $\operatorname{Ker}\Phi \cap \operatorname{Ker}\mathrm{D}^2 J_T(x^\star) = \{0\}$.

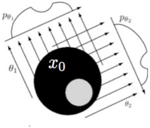
$J = \|\cdot\|_1 \to df(Y) = |\operatorname{supp}(x^\star(Y))|$

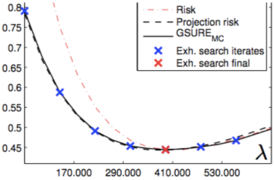
Projected GSURE: unbiased estimate of $\|P^\perp_{\operatorname{Ker}\Phi}(x^\star(Y) - x_0)\|^2$

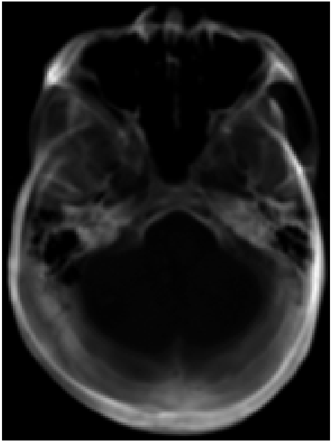
An Application to Tomography

Φ sub-sampled Radon transform (16)

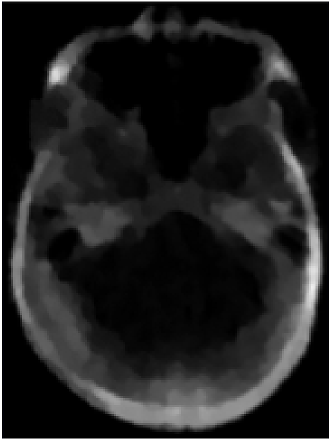
$J = \|\nabla_d \cdot\|_{\mathcal{B}}$







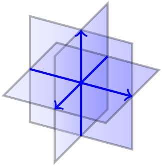
x_0



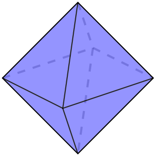
$x^\star_{\lambda^\star}$

Summary

Partial smoothness: encodes models using singularities



$\longrightarrow$



Performance

- $\odot$ ℓ^2 error
- $\odot$ model stability

Parameter selection

- $\odot$ local behavior
- $\odot$ unbiased risk estimation

$\Longrightarrow$ **Sensitivity analysis**

Future Work

Extended-valued functions: minimization under constraints

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \quad \text{subject to} \quad x \geqslant 0$$

Future Work

Extended-valued functions: minimization under constraints

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \quad \text{subject to} \quad x \geqslant 0$$

Non-convexity: Fidelity and regularization, dictionary learning

$$\min_{x_k \in \mathbb{R}^n, D \in \mathcal{D}} \sum_k \frac{1}{2}\|y - \Phi D x_k\|^2 + \lambda J(x_k)$$

Future Work

Extended-valued functions: minimization under constraints

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \quad \text{subject to} \quad x \geqslant 0$$

Non-convexity: Fidelity and regularization, dictionary learning

$$\min_{x_k \in \mathbb{R}^n, D \in \mathcal{D}} \sum_k \frac{1}{2}\|y - \Phi D x_k\|^2 + \lambda J(x_k)$$

Infinite dimensional problems: partial smoothness for BV, Besov

$$\min_{f \in \mathrm{BV}(\Omega) \cap \mathrm{L}^2(\Omega)} \frac{1}{2}\|g - \Psi f\|_{\mathrm{L}^2(\Omega)} + \lambda |\mathrm{D}f|(\Omega)$$

Future Work

Extended-valued functions: minimization under constraints

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x) \quad \text{subject to} \quad x \geqslant 0$$

Non-convexity: Fidelity and regularization, dictionary learning

$$\min_{x_k \in \mathbb{R}^n, D \in \mathcal{D}} \sum_k \frac{1}{2}\|y - \Phi D x_k\|^2 + \lambda J(x_k)$$

Infinite dimensional problems: partial smoothness for BV, Besov

$$\min_{f \in \mathrm{BV}(\Omega) \cap \mathrm{L}^2(\Omega)} \frac{1}{2}\|g - \Psi f\|_{\mathrm{L}^2(\Omega)} + \lambda |\mathrm{D} f|(\Omega)$$

Compressed sensing: Optimal bounds for partly smooth regularizers