Iterative regularization via proximal methods

Silvia Villa

Dipartimento di Matematica
Politecnico di Milano

CosmoStat Day on Machine Learning in Astrophysics
Paris, January 26th, 2018

Outline

Introduction and motivation

Outline

Introduction and motivation

Part I: Quadratic data fit
Joint work with: **S.Matet, L. Rosasco, B.C. Vũ**

Part II: General data fit
Joint work with: **G. Garrigos, L. Rosasco**

Computational regularization for large scale data problems

Integrating

REGULARIZATION and **OPTIMIZATION**

in inverse problems (and learning)

Computational regularization for large scale data problems

Integrating

REGULARIZATION and **OPTIMIZATION**

in inverse problems (and learning)

Computational requirements tailored to the information in the data rather than to their raw amount

Inverse problems

- H and G Hilbert spaces
- $A\colon H \to G$ linear and bounded

Goal

Let $y \in G$, approximate the solution of

$$Ax = y,$$

assuming that a solution exists.

Inverse problems

- H and G Hilbert spaces
- $A\colon H \to G$ linear and bounded

Goal

Let $y \in G$, approximate the solution of

$$Ax = y,$$

assuming that a solution exists.

Selection principle: given $R\colon H \to \mathbb{R} \cup \{+\infty\}$ strongly convex, select $x^\dagger$, the solution of

$$\begin{aligned} &\min R(x) \\ &\text{s.t. } Ax = y \end{aligned}$$

Noisy data

We do not know $y \in \mathcal{G}$. We have access only to $\widehat{y} \in \mathcal{G}$ such that

$$\|\widehat{y} - y\| \leq \delta, \qquad \delta > 0.$$

Goal:

find a **stable** approximation of $x^\dagger$

using only $\widehat{y}$.

Tikhonov regularization

Consider the regularized problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \frac{1}{2\lambda}\|Ax - \widehat{y}\|^2 + R(x)$$

Tikhonov regularization

Consider the regularized problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \frac{1}{2\lambda}\|Ax - \widehat{y}\|^2 + R(x)$$

How to choose λ*?*

Tikhonov regularization

Consider the regularized problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \ \ \frac{1}{2\lambda}\|Ax-\widehat{y}\|^2 + R(x)$$

How to choose λ?

Well-established regularization theory

Theorem

Let $\widehat{x}^{\lambda}$ be the solution of the regularized problem. Then

$$\|\widehat{x}^{\lambda} - x^{\dagger}\| \leq C\left(\frac{\delta}{\sqrt{\lambda}} + \sqrt{\delta} + \sqrt{\lambda}\right)$$

Choosing $\lambda_\delta \sim \delta$, we derive

$$\|\widehat{x}^{\lambda_\delta} - x^{\dagger}\| \leq C\sqrt{\delta}.$$

[Burger-Osher, Convergence rates of convex variational regularization, 2004]

Tikhonov regularization

What about computations?

Tikhonov regularization

What about computations?

Tikhonov regularization in practice:

- choose an interval $[\lambda_{\min}, \lambda_{\max}]$
- approximately solve the regularized problem for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$
- select the best λ according to a validation criterion

Tikhonov regularization

What about computations?

Tikhonov regularization in practice:

- choose an interval $[\lambda_{\min}, \lambda_{\max}]$
- approximately solve the regularized problem for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$
- select the best λ according to a validation criterion

Theorem

Let $\widehat{x}^{\lambda,\varepsilon}$ be an ε-solution of the relaxed problem. Then choosing

$$\|\widehat{x}^{\lambda,\varepsilon} - x^\dagger\| \leq C \left(\frac{\delta + \sqrt{\varepsilon}}{\sqrt{\lambda}} + \sqrt{\delta} + \sqrt{\lambda} \right).$$

Choosing $\lambda_\delta \sim \delta$ and $\varepsilon_\delta \sim \delta^2$, we derive

$$\|\widehat{x}^{\lambda_\delta,\varepsilon_\delta} - x^\dagger\| \leq C\sqrt{\delta}.$$

Iterative regularization

A (new) old idea

Solve:

$$\min_{Ax=y} R(x)$$

.

Iterative regularization

A (new) old idea

Solve:

$$\min_{Ax=\widehat{\mathbf{y}}} R(x)$$

and **early stop the iterations**.

Iterative regularization

A (new) old idea

Solve:

$$\min_{Ax=\widehat{\mathbf{y}}} R(x)$$

and **early stop the iterations**.
An old idea in inverse problems for $R = \|\cdot\|^2/2$:
Landweber [Engl-Hanke-Neubauer, inverse problems]
Recently revisited: [Osher-Burger-Yin-Cai-Resmerita-He.....$\sim$ 2000s]

Iterative regularization: idea of the proof

1. Choose a convergent algorithm to solve

$$\min_{Ax=y} R(x).$$

   Call the iterates $(x_t)_{t\in\mathbb{N}}$.

2. Apply the same algorithm to

$$\min_{Ax=\widehat{y}} R(x).$$

   Call the iterates $(\widehat{x}_t)_{t\in\mathbb{N}}$.

3. Then

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

Iterative regularization: idea of the proof

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

$$x_t \longrightarrow x_{t+1} \qquad \longrightarrow \ldots$$

Iterative regularization: idea of the proof

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

$$x_t \longrightarrow x_{t+1} \quad \longrightarrow \ldots \quad \longrightarrow x^\dagger$$

Iterative regularization: idea of the proof

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

$$x_t \longrightarrow x_{t+1} \quad \longrightarrow \dots \quad \longrightarrow x^\dagger$$

$$\widehat{x}_t$$

$$\searrow$$

Iterative regularization: idea of the proof

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

$$x_t \longrightarrow x_{t+1} \longrightarrow \dots \longrightarrow x^\dagger$$

$\widehat{x}_t$

$\searrow$

$\widehat{x}_{t+1}$

$\searrow$

$\dots$

$\searrow$

a solution of the noisy problem

(if it exists)

Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image $\hat{x}_t$ $\hat{y}$

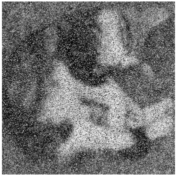
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image $\hat{x}_t$ $\hat{y}$

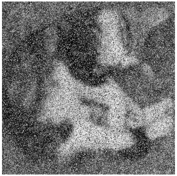
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

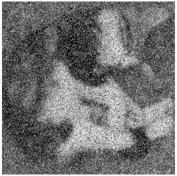
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

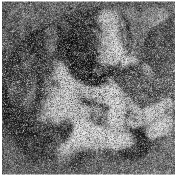
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

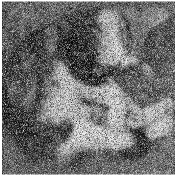
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \le \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

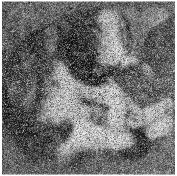
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$


original image


$\hat{x}_t$


$\hat{y}$

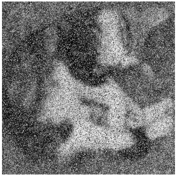
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

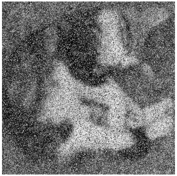
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



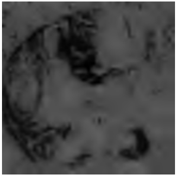
$\hat{y}$

Iterative regularization at work

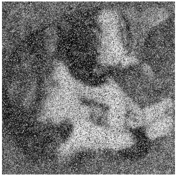
Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image



$\hat{x}_t$

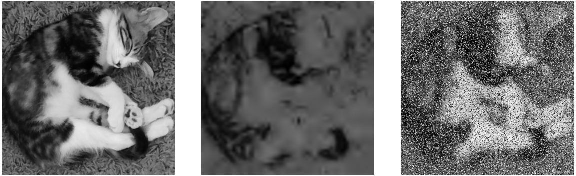


$\hat{y}$

Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

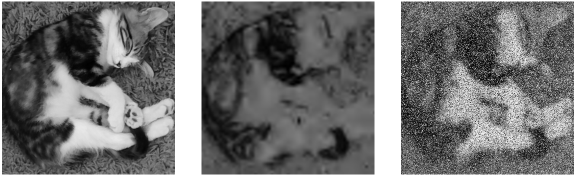


original image $\hat{x}_t$ $\hat{y}$

Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

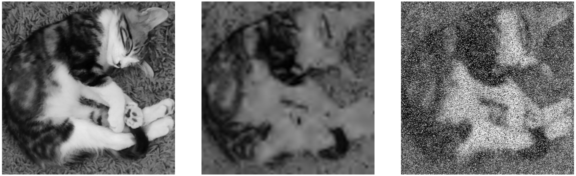


original image $\hat{x}_t$ $\hat{y}$

Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$



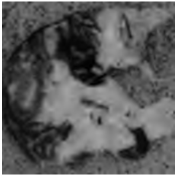
original image $\hat{x}_t$ $\hat{y}$

Iterative regularization at work

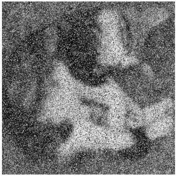
Recall that

$$\|\widehat{x}_t - x^\dagger\| \le \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image



$\hat{x}_t$



$\hat{y}$

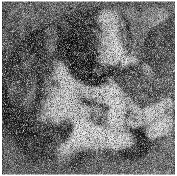
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

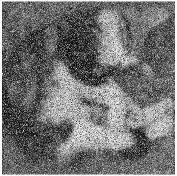
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

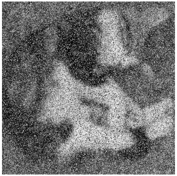
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

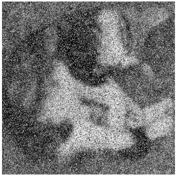
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

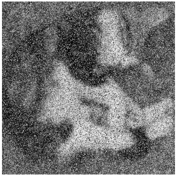
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

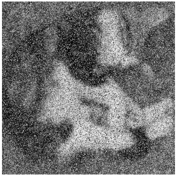
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \le \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

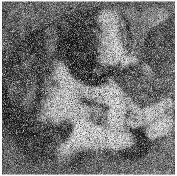
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

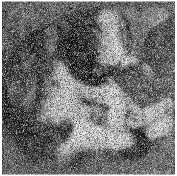
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

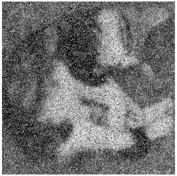
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$


original image


$\hat{x}_t$


$\hat{y}$

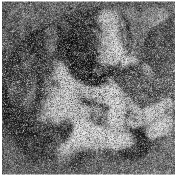
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

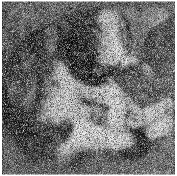
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

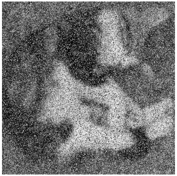
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

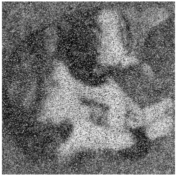
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

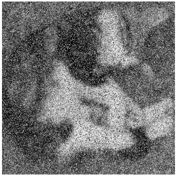
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

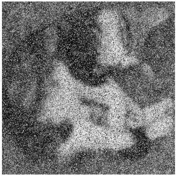
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

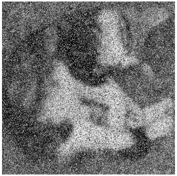
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

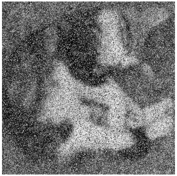
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

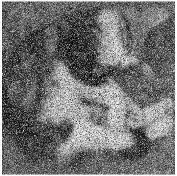
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

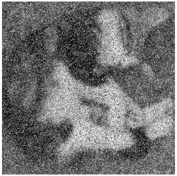
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

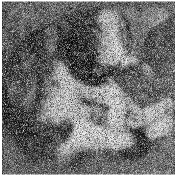
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

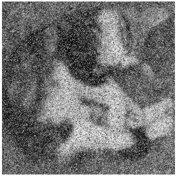
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



$\hat{y}$

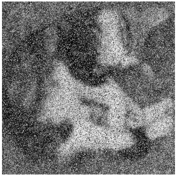
Iterative regularization at work

Recall that

$$\|\widehat{x}_t - x^{\dagger}\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^{\dagger}\|}_{\text{optimization}}$$

original image

$\hat{x}_t$



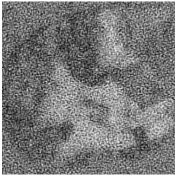
$\hat{y}$

Iterative regularization at work

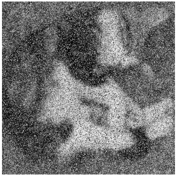
Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image



$\hat{x}_t$



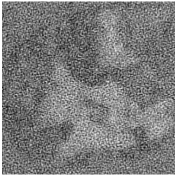
$\hat{y}$

Iterative regularization at work

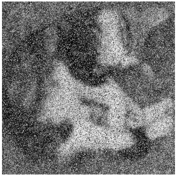
Recall that

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

original image



$\hat{x}_t$



$\hat{y}$

Dual problem - exact datum

$$\min_{Ax=y} R(x) \quad \longleftrightarrow \quad \min_{w\in\mathcal{H}} R(x) + \iota_{\{y\}}(Ax),$$

where $\iota_{\{y\}}(u) = 0$ if $u = y$ and $\iota_{\{y\}}(u) = +\infty$ otherwise.

Dual problem - exact datum

$$\min_{Ax=y} R(x) \quad \longleftrightarrow \quad \min_{w\in\mathcal{H}} R(x) + \iota_{\{y\}}(Ax),$$

where $\iota_{\{y\}}(u) = 0$ if $u = y$ and $\iota_{\{y\}}(u) = +\infty$ otherwise.

The dual problem is

$$\min_{v\in\mathcal{G}} d(v), \quad d(v) = R^*(-A^*v) + \langle y, v\rangle.$$

R strongly convex $\Rightarrow$ **the dual is smooth**

We can apply gradient method to it, or an **accelerated** gradient method.

Dual gradient descent

R strongly convex $\Rightarrow$

$$R = F + \frac{\alpha}{2}\|\cdot\|^2 \quad \text{for some convex function } F.$$

Let $v_0 \in \mathcal{G}$, and let $\gamma \in \left]0, \alpha\|A\|^{-2}\right[$. Iterate

$$\begin{aligned} v_{t+1} &= v_t - \gamma(\nabla(R^* \circ -A^*)(v_t) + y) \\ &= v_t + \gamma(A\,\mathrm{prox}_{\alpha^{-1}F}(-\alpha^{-1}A^* v_t) - y) \end{aligned}$$

Dual gradient descent

R strongly convex $\Rightarrow$

$$R = F + \frac{\alpha}{2}\|\cdot\|^2 \quad \text{for some convex function } F.$$

Let $v_0 \in \mathcal{G}$, and let $\gamma \in \left]0, \alpha\|A\|^{-2}\right[$. Iterate

$$\begin{aligned} v_{t+1} &= v_t - \gamma(\nabla(R^* \circ -A^*)(v_t) + y) \\ &= v_t + \gamma(A \underbrace{\operatorname{prox}_{\alpha^{-1}F}(-\alpha^{-1}A^* v_t)}_{=x_t} - y) \end{aligned}$$

Equivalent to:

$$\begin{cases} x_t = \operatorname{prox}_{\alpha^{-1}F}(-\alpha^{-1}A^* v_t) \\ v_{t+1} = v_t + \gamma(Ax_t - y) \end{cases}$$

A.k.a. linearized Bregman iteration [Yin-Osher-Burger, several papers, Bachmayr-Burger, 2005]

Landweber algorithm

When $R = \|\cdot\|^2/2$, then $F = 0$, and

$$\begin{aligned} x_t &= -A^* v_t \\ x_{t+1} &= x_t - \gamma A^*(Ax_t - y) \\ &= (I - \gamma A^* A)x_t + \gamma A^* y \end{aligned}$$

Landweber algorithm

When $R = \|\cdot\|^2/2$, then $F = 0$, and

$$\begin{aligned} x_t &= -A^* v_t \\ x_{t+1} &= x_t - \gamma A^*(Ax_t - y) \\ &= (I - \gamma A^* A)x_t + \gamma A^* y \end{aligned}$$

Gradient method applied to $(1/2)\|Ax - y\|^2$

Accelerated variant: FISTA on the dual

Let $v_0 = z_{-1} = z_0 = 0 \in \mathcal{G}$, $\gamma = \alpha\|A\|^{-2}$, $\theta_0 = 1$ and $\theta_{t+1} = (1 + \sqrt{1 + 4\theta_t^2})/2$.

$$
\begin{cases}
x_t = \operatorname{prox}_{\alpha^{-1}F}(-\alpha^{-1}A^* z_t) \\
r_t = \operatorname{prox}_{\alpha^{-1}F}\big(-\alpha^{-1}A^* v_t\big) \\
z_{t+1} = v_t + \gamma(Ar_t - y) \\
v_{t+1} = z_{t+1} + \frac{\theta_t - 1}{\theta_{t+1}}(z_{t+1} - z_t)
\end{cases}
$$

Accelerated Landweber algorithm

When $R = \|\cdot\|^2/2$, then $x_t = -A^* v_t$ and $r_t = -A^* z_t$,

$$\begin{cases} x_{t+1} = x_t - \gamma A^*(Ar_t - y) \\ r_{t+1} = x_{t+1} + \frac{\theta_t - 1}{\theta_{t+1}}(x_{t+1} - x_t) \end{cases}$$

Accelerated Landweber algorithm

When $R = \|\cdot\|^2/2$, then $x_t = -A^* v_t$ and $r_t = -A^* z_t$,

$$\begin{cases} x_{t+1} = x_t - \gamma A^*(A r_t - y) \\ r_{t+1} = x_{t+1} + \frac{\theta_t - 1}{\theta_{t+1}}(x_{t+1} - x_t) \end{cases}$$

FISTA applied to $(1/2)\|Ax - y\|^2$

A technical condition

1. Existence of the solution of the dual (for the exact y) needed for convergence rates
2. From convergence on the dual to convergence on the primal

Qualification (source) condition (Only for the exact datum)

There exists $q \in \mathcal{G}$ such that

$$A^* q \in \partial R(x^\dagger)$$

A technical condition

1. Existence of the solution of the dual (for the exact y) needed for convergence rates
2. From convergence on the dual to convergence on the primal

Qualification (source) condition (Only for the exact datum)

There exists $q \in \mathcal{G}$ such that

$$A^* q \in \partial R(x^\dagger)$$

Same condition needed for establishing rates for Tikhonov regularization.

Convergence for dual gradient descent

Theorem

Assume qualification condition. Let $v^\dagger$ be a solution of the dual problem. For every $t \in \mathbb{N}$

$$\|x_t - x^\dagger\| \leq \frac{2}{\alpha}(d(v_t) - d(v^\dagger))^{1/2} \leq \frac{\|A\|\|v_0 - v^\dagger\|}{\alpha\sqrt{t}}$$

Stability

Theorem (Matet-Rosasco-V.-Vu, 2017)

There exists $t_\delta \sim \delta^{-1}$ *such that*

$$\|\widehat{x}_t - x_t\| \leq \frac{2}{\|A\|}\sqrt{t}\delta.$$

Iterative regularization result

How to choose the stopping time?

Theorem (Dual gradient descent)

There exists $t_\delta \sim \delta^{-1}$ such that

$$\|\widehat{x}_{t_\delta} - x^\dagger\| \leq C\delta^{1/2}.$$

Accelerated dual gradient descent

How to choose the stopping time?

Theorem (Accelerated Dual gradient descent)

Assume the qualification condition. Then, if $t_\delta \sim \delta^{-1/2}$,

$$\|\widehat{x}_{t_\delta} - x^\dagger\| \leq C\delta^{1/2}.$$

Based on the results of [Aujol-Dossal, 2016]

Accelerated dual gradient descent

How to choose the stopping time?

Theorem (Accelerated Dual gradient descent)

Assume the qualification condition. Then, if $t_\delta \sim \delta^{-1/2}$,

$$\|\widehat{x}_{t_\delta} - x^\dagger\| \leq C\delta^{1/2}.$$

Based on the results of [Aujol-Dossal, 2016]
For $R = \|\cdot\|^2/2$ see also [A. Neubauer, On Nesterov acceleration for Landweber iteration of linear ill-posed problems, Nov. 2016]

Back to the beginning: regularized inverse problems

Tikhonov regularization: original hierarchical problem is replaced by

$$\text{minimize} \quad \frac{1}{\lambda} D(Ax, y) + R(x),$$

for a suitable $\lambda > 0$, and an algorithm is chosen to compute

$$x_{t+1} = \text{Algo}(x_t, \lambda).$$

Back to the beginning: regularized inverse problems

Tikhonov regularization: original hierarchical problem is replaced by

$$\text{minimize} \quad \frac{1}{\lambda} D(Ax, y) + R(x),$$

for a suitable $\lambda > 0$, and an algorithm is chosen to compute

$$x_{t+1} = \text{Algo}(x_t, \lambda).$$

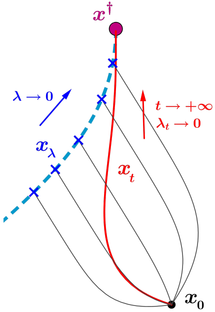
A diagonal approach[Lemaire 80s-90s]

$$x_{t+1} = \text{Algo}(x_t, \lambda_t),$$

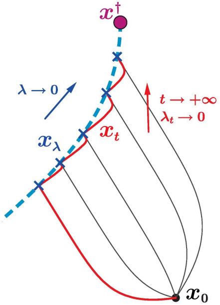
with $\lambda_t \to 0$.

A picture

The previous approach allows to describe:



A diagonal strategy



A warm restart strategy

A dual approach

Diagonal forward-backward: [Attouch, Cabot, Czarnecki, Peypouquet ...]

A dual approach

Diagonal forward-backward: [Attouch, Cabot, Czarnecki, Peypouquet ...]
Not well-suited if D is not smooth.

A dual approach

Diagonal forward-backward: [Attouch, Cabot, Czarnecki, Peypouquet ...]
Not well-suited if D is not smooth.

$$\begin{array}{ccc} \min R(x) \\ \text{s.t. } D(Ax,y)=0 & \longrightarrow & \frac{1}{\lambda}D(Ax,y)+R(x) \\ \uparrow & & \downarrow \\ \min_{u\in G} \underbrace{\langle u,y\rangle + R^*(-A^*u)}_{=d(u)} & \longleftarrow & \underbrace{\frac{1}{\lambda}D^*(\lambda u,y)+R^*(-A^*u)}_{=d_\lambda(u)}. \end{array}$$

Dual diagonal descent algorithm (3D)

If $R = F + (\sigma_R/2)\|\cdot\|^2$ is strongly convex:

$$d_\lambda(u) = \underbrace{R^*(-A^*u)}_{\text{smooth}} + \underbrace{\frac{1}{\lambda}D^*(\lambda u, y)}_{\text{nonsmooth}}$$

We can use the **forward-backward splitting algorithm** on the dual.

$$\left| \begin{array}{l} u_0 \in G, \ \lambda_t \to 0, \tau = \sigma_R/\|A\|^2 \\ \\ z_{t+1} = u_t + \tau A \nabla R^*(-A^* u_t) \\ u_{t+1} = \mathrm{prox}_{\tau \lambda_t^{-1} D^*(\lambda_t \cdot, y)}(z_{t+1}). \end{array} \right.$$

Dual diagonal descent algorithm (3D)

If $R = F + (\sigma_R/2)\|\cdot\|^2$ is strongly convex:

$$d_\lambda(u) = \underbrace{R^*(-A^* u)}_{\text{smooth}} + \underbrace{\frac{1}{\lambda} D^*(\lambda u, y)}_{\text{nonsmooth}}$$

We can use the **forward-backward splitting algorithm** on the dual.

$$u_0 \in G, \ \lambda_{\mathbf{t}} \to \mathbf{0}, \tau = \sigma_R/\|A\|^2$$

$$z_{t+1} = u_t + \tau A \nabla R^*(-A^* u_t)$$

$$u_{t+1} = z_{t+1} - \tau \operatorname{prox}_{(\tau\lambda_t)^{-1} D(\cdot, y)} \left(\tau^{-1} z_{t+1}\right)$$

Dual diagonal descent algorithm (3D)

If $R = F + (\sigma_R/2)\|\cdot\|^2$ is strongly convex:

$$d_\lambda(u) = \underbrace{R^*(-A^*u)}_{\text{smooth}} + \underbrace{\frac{1}{\lambda}D^*(\lambda u, y)}_{\text{nonsmooth}}$$

We can use the **forward-backward splitting algorithm** on the dual.

$$\left|\begin{array}{l} u_0 \in G,\ \lambda_{\mathbf{t}} \to \mathbf{0}, \tau = \sigma_R/\|A\|^2 \\ x_t = \nabla R^*(-A^*u_t) = \operatorname{prox}_{\sigma_R^{-1}F}(-A^*u_t) \\ z_{t+1} = u_t + \tau A x_t \\ u_{t+1} = z_{t+1} - \tau \operatorname{prox}_{(\tau\lambda_t)^{-1}D(\cdot,y)}\left(\tau^{-1}z_{t+1}\right) \end{array}\right.$$

Convergence of diagonal dual descent algorithm

AD1) $D\colon G \times G \to [0, +\infty]$ and $D(u, y) = 0 \iff u = y$.

Convergence of diagonal dual descent algorithm

AD1) $D\colon G \times G \to [0, +\infty]$ and $D(u, y) = 0 \iff u = y$.

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p-well conditioned

Convergence of diagonal dual descent algorithm

AD1) $D\colon G \times G \to [0, +\infty]$ and $D(u, y) = 0 \iff u = y$.

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p-well conditioned

AR) There exists a solution $\bar{x}$ such that $A\bar{x} = y \qquad \bar{x} \in \operatorname{dom} R$.

Convergence of diagonal dual descent algorithm

AD1) $D\colon G \times G \to [0, +\infty]$ and $D(u, y) = 0 \iff u = y$.

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p-well conditioned

AR) There exists a solution $\bar{x}$ such that $A\bar{x} = y \qquad \bar{x} \in \operatorname{dom} R$.

Theorem [Garrigos-Rosasco-V. 2017]

Suppose that $\lambda_t \in \ell^{1/(p-1)}(\mathbb{N})$. Let $x^\dagger$ be the solution of (P). Assume that there exists $q \in \mathcal{G}$ such that

$$A^* q \in \partial R(x^\dagger)$$

Then $\|x_t - x^\dagger\| = o(t^{-1/2})$

Stability

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

Stability

$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

Stability Theorem [Garrigos-Rosasco-V. 2017]

Assume that the source/qualification condition holds. Let $\hat{y} \in Y$, with $\|\hat{y} - y\| \leq \delta$. Let $(\hat{x}_t, \hat{u}_t)$ be the sequence generated by the (3D) algorithm with $y = \hat{y}$ and $\hat{u}_0 = u_0$. Suppose that

$$(\lambda_t) \in \ell^{1/(p-1)}(\mathbb{N}).$$

Then

$$\|x_t - \hat{x}_t\| \leq C\delta t.$$

For simplicity here $D(u, y) = L(u - y)$. But this is not needed.

Stability with respect to errors = iterative regularization results

Theorem (Early-stopping) [Garrigos-Rosasco-V. 2017]

Assume that the source/qualification condition holds. Let $\hat{y} \in Y$, with $\|\hat{y} - y\| \leq \delta$. Let $(\hat{x}_t, \hat{u}_t)$ be the sequence generated by the (3D) algorithm with $y = \hat{y}$ and $\hat{u}_0 = u_0$. Suppose that

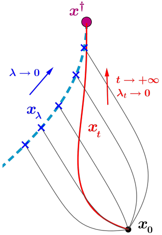
$$(\lambda_t) \in \ell^{1/(p-1)}(\mathbb{N})$$

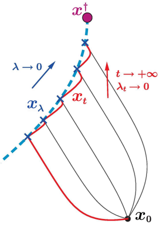
Then there exists an early stopping rule $t(\delta) = \lceil c\delta^{-2/3} \rceil$ which verifies

$$\|\hat{x}_{t(\delta)} - x^\dagger\| = O(\delta^{\frac{1}{3}}) \text{ when } \delta \to 0.$$

Setting

- deblurring and denoising (salt and pepper, gaussian, gaussian+salt and pepper, Poisson) of 512 x 512 images
- comparison between the two versions: **diagonal** and **warm restart**



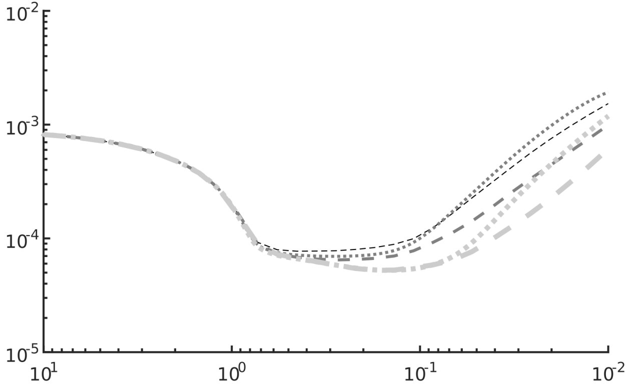


diagonal:
one parameter = (λ_t)= n. iter.

warm restart:
two parameters: (λ_t); accuracy

Diagonal works as well as warm restart (i.e. Tikhonov)

Euclidean distance from the true image

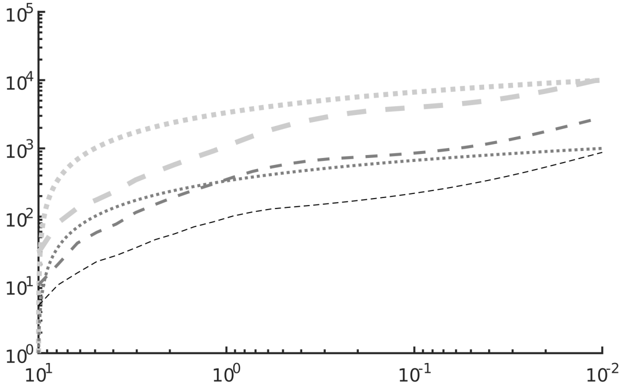


Dotted lines: diagonal with 10^3 and 10^4 iterations

Dashed lines: warm restart with 30 λs and accuracy : $10^{-3}, 10^{-4}, 10^{-5}$

Diagonal works better than(?) warm restart (i.e. Tikhonov)

Total number of iterations as a function of (λ_t)



Dotted lines: diagonal

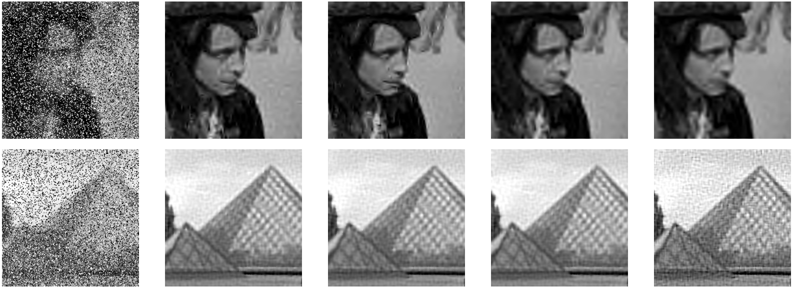
Dashed lines: warm restart with 30 λs and accuracy: $10^{-3}, 10^{-4}, 10^{-5}$

Parameter selection

- using the true image
- using SURE (and the ideas in : Deladalle-Vaiter-Fadili-Peyré 2014 to compute it)
- budget of 10^3 iterations for diagonal and warm restart

Results

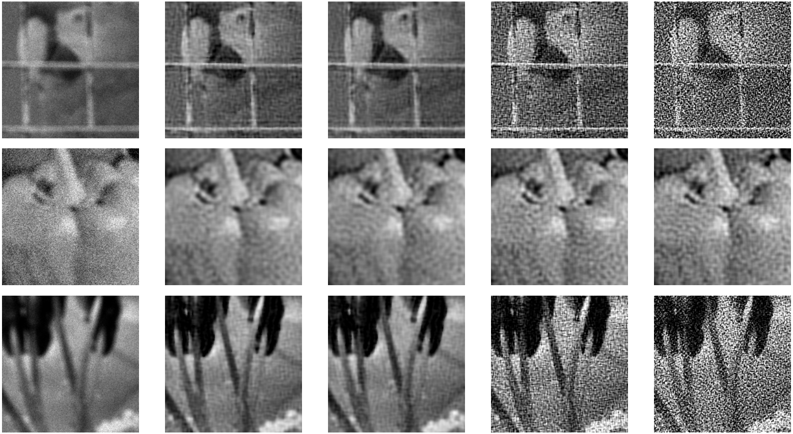
Blurring + Salt and pepper 35%. $D(u, y) = \|u - y\|_1$,
$R(x) = \|Wx\|_1 + \|x\|^2$ or $\|x\|_{TV} + \|x\|^2$



noisy image, reconstruction with diagonal and warm restart using true image, reconstruction with diagonal and warm restart using SURE

Results

Blurring + Poisson noise. $D(u, y) = \mathrm{KL}(y; u + b)$, $R(x) = \|x\|_{TV} + \|x\|^2$



Concluding remarks ad future perspecitves

Concluding remarks

- use the number of iterations as regularization parameters
- iterative regularization as an alternative to Tikhonov regularization
- optimization perspective: stability with respect to errors as a way to prove regularization results

Concluding remarks ad future perspecitves

Concluding remarks

- use the number of iterations as regularization parameters
- iterative regularization as an alternative to Tikhonov regularization
- optimization perspective: stability with respect to errors as a way to prove regularization results

Future perspectives

- accelerated version of diagonal Tikhonov
- remove strong convexity
- better use of conditioning?

References

- S. Matet, L. Rosasco, B. C. Vũ, Don't relax: early stopping for convex regularization, arxiv 2016.
- G. Garrigos, L. Rosasco, and S. Villa, Iterative regularization via dual diagonal descent, JMIV 2017

The end

Merci pour votre attention