Stability and regularization properties of diagonal proximal gradient methods

Silvia Villa

Department of Mathematics, MaLGa Center
University of Genoa, https://ml.unige.it

Workshop “Regularisation for inverse problems”, Paris, November 19th, 2019

UniGe | MaLGa

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

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

Joint work with: **G. Garrigos, L. Rosasco and L. Calatroni**

Inverse problems

- H and G Hilbert spaces.
- $A \in \mathcal{L}(H, G)$ forward operator

Inverse problem

Given $y \in G$, find $x \in H$ s.t. $Ax = y$

Inverse problems

- H and G Hilbert spaces.
- $A \in \mathcal{L}(H, G)$ forward operator
- **Ill-posedness!** (existence? uniqueness? stability?)

Inverse problems in practice

Given $y \in G$, how to solve $Ax = y$?

Inverse problems

- H and G Hilbert spaces.
- $A \in \mathcal{L}(H, G)$ forward operator
- **Ill-posedness!** (existence? uniqueness? stability?)

Inverse problems in practice

Given $y \in G$, how to solve $Ax = y$?

- **EXISTENCE**: introduce data discrepancy

$$x^\dagger = \operatorname*{argmin}_x \; D(Ax, y)$$

Inverse problems

- H and G Hilbert spaces.
- $A \in \mathcal{L}(H, G)$ forward operator
- **Ill-posedness!** (existence? uniqueness? stability?)

Inverse problems in practice

Given $y \in G$, how to solve $Ax = y$?

- **EXISTENCE**: introduce data discrepancy

$$x^\dagger = \underset{x}{\operatorname{argmin}} \; D(Ax, y)$$

- **UNIQUENESS**: introduce *a-priori* information on x. Let $R : H \to R \cup \{+\infty\}$ be strongly convex, and define

$$x^\dagger = \underset{x \in \operatorname{argmin} D(Ax,y)}{\operatorname{argmin}} \; R(x)$$

UniGe | MaLGa

Inverse problems

- H and G Hilbert spaces.
- $A \in \mathcal{L}(H, G)$ forward operator
- **Ill-posedness!** (existence? uniqueness? stability?)

Inverse problems in practice

Given $y \in G$, how to solve $Ax = y$?

- **EXISTENCE**: introduce data discrepancy

$$x^\dagger = \underset{x}{\operatorname{argmin}} \; D(Ax, y)$$

- **UNIQUENESS**: introduce *a-priori* information on x. Let $R : H \to R \cup \{+\infty\}$ be strongly convex, and define

$$x^\dagger = \underset{x \in \operatorname{argmin} D(Ax,y)}{\operatorname{argmin}} \; R(x)$$

- **STABILITY**: perturbation of the data...

Data perturbations & stability

In practice, we do not have access to $y \in G$, but only to its **noisy** version $\widehat{y} \in G$:

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

Goal

Given $\widehat{y}$, find a **stable** approximation of $x^\dagger$.

UniGe | MaLGa

Data perturbations & stability

In practice, we do not have access to $y \in G$, but only to its **noisy** version $\widehat{y} \in G$:

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

Goal

Given $\widehat{y}$, find a **stable** approximation of $x^\dagger$.

$$x^\dagger := \operatorname*{argmin}_{x \in \operatorname{argmin} D(Ax,y)} R(x)$$

$x^\dagger$

$y = Ax^\dagger$

UniGe | MaLGa

Data perturbations & stability

In practice, we do not have access to $y \in G$, but only to its **noisy** version $\widehat{y} \in G$:

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

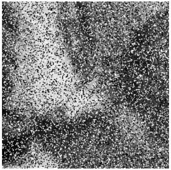
Goal

Given $\widehat{y}$, find a **stable** approximation of $x^\dagger$.

$$\widehat{x}^\dagger := \operatorname*{argmin}_{x \in \operatorname{argmin} D(Ax, \widehat{y})} R(x),$$

$x^\dagger$

$y = Ax^\dagger$



$\widehat{y}$

UniGe | MaLGa

Data perturbations & stability

In practice, we do not have access to $y \in G$, but only to its **noisy** version $\widehat{y} \in G$:

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

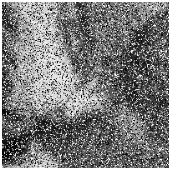
Goal

Given $\widehat{y}$, find a **stable** approximation of $x^\dagger$.

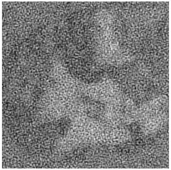
$$\widehat{x}^\dagger := \operatorname*{argmin}_{x \in \operatorname{argmin} {\color{blue}D(Ax,\widehat{y})}} {\color{red}R(x)},$$

$x^\dagger$

$y = Ax^\dagger$



$\widehat{y}$



$\widehat{x}^\dagger$

Data perturbations & stability

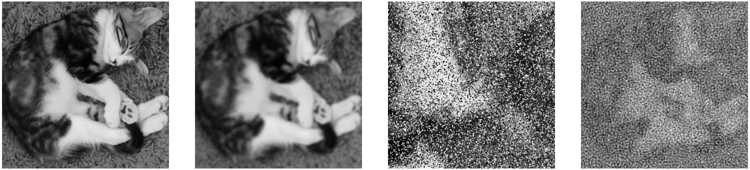
In practice, we do not have access to $y \in G$, but only to its **noisy** version $\widehat{y} \in G$:

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

Goal

Given $\widehat{y}$, find a **stable** approximation of $x^\dagger$.

$$\widehat{x}^\dagger := \operatorname*{argmin}_{x \in \operatorname{argmin} D(Ax, \widehat{y})} R(x),$$



$x^\dagger$ $\quad$ $y = Ax^\dagger$ $\quad$ $\widehat{y}$ $\quad$ $\widehat{x}^\dagger$

How to enforce well-posedness?

Tikhonov regularization

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

Tikhonov regularization

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

How to choose λ?

UniGe | MaLGa

Tikhonov regularization

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \frac{1}{2\lambda} D(Ax,\widehat{y}) + R(x)$$

How to choose λ?

Theorem

Let $D(Ax,y) = \|Ax-y\|^2$. Let $\widehat{x}^\lambda$ be the solution of the regularized problem and assume that $Im(A^*)\cap\partial R(x^\dagger)\neq\varnothing$. 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]

[Benning-Burger, Error estimates for general fidelities, 2011]

UniGe | MaLGa

Tikhonov regularization

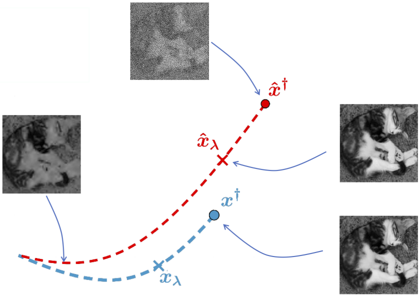
What about computations?

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



Iterative regularization

A (new) old idea

Solve:

$$\min_{x \in \operatorname{argmin} D(A\cdot, \widehat{y})} R(x)$$

.

Iterative regularization

A (new) old idea

Solve:

$$\min_{x \in \operatorname{argmin} D(A\cdot, \widehat{y})} R(x)$$

BUT early stop the iterations.

Iterative regularization

A (new) old idea

Solve:

$$\min_{x \in \operatorname{argmin} D(A\cdot, \widehat{y})} R(x)$$

BUT 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 find

$$x^\dagger \in \operatorname*{argmin}_{x \in \operatorname{argmin} D(A\cdot, y)} R(x)$$

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

2. Apply the same algorithm to

$$\operatorname*{argmin}_{x \in \operatorname{argmin} D(A\cdot, \widehat{y})} R(x)$$

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

3. Then

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

UniGe | MaLGa

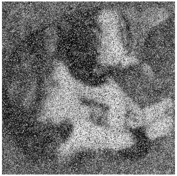
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

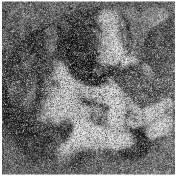
Iterative regularization at work

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


$\hat{x}_t$


$\hat{y}$

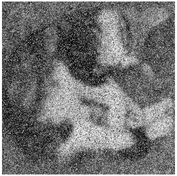
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

UniGe | MaLGa

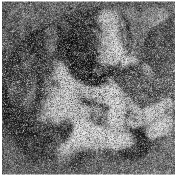
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

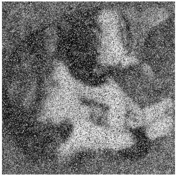
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

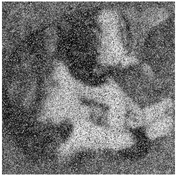
Iterative regularization at work

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

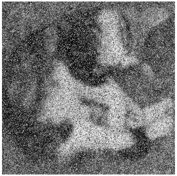
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

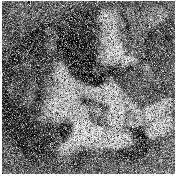
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

UniGe | MaLGa

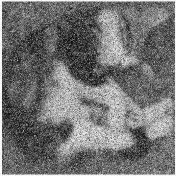
Iterative regularization at work

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

$\hat{x}_t$



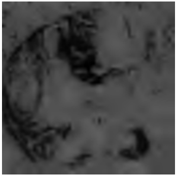
$\hat{y}$

Iterative regularization at work

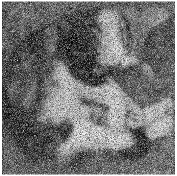
Recall that

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

original image



$\hat{x}_t$



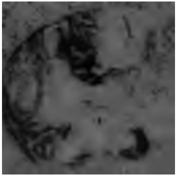
$\hat{y}$

Iterative regularization at work

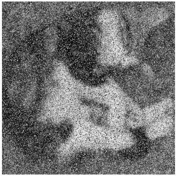
Recall that

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

original image



$\hat{x}_t$



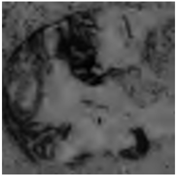
$\hat{y}$

Iterative regularization at work

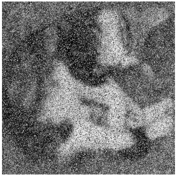
Recall that

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

original image



$\hat{x}_t$



$\hat{y}$

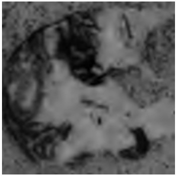
UniGe | MaLGa

Iterative regularization at work

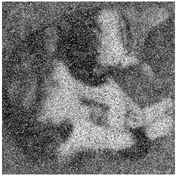
Recall that

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

original image



$\hat{x}_t$



$\hat{y}$

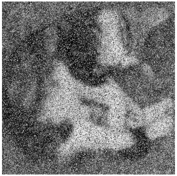
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

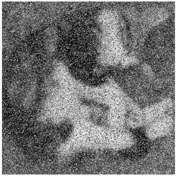
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

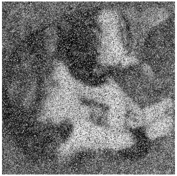
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

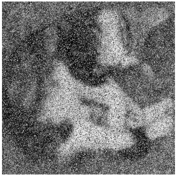
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

UniGe | MaLGa

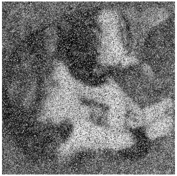
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

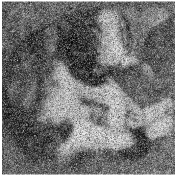
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

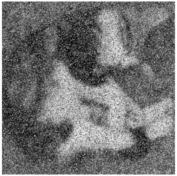
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

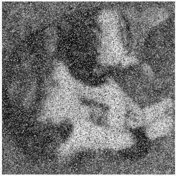
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

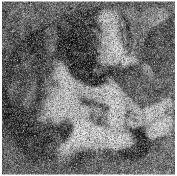
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

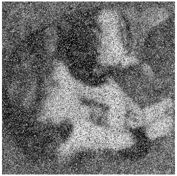
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

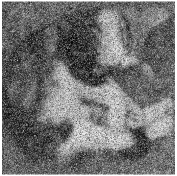
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

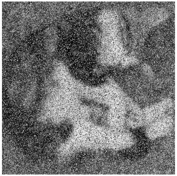
Iterative regularization at work

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


$\hat{x}_t$


$\hat{y}$

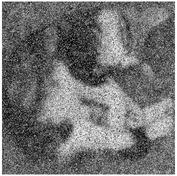
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

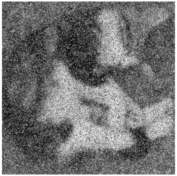
Iterative regularization at work

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


$\hat{x}_t$


$\hat{y}$

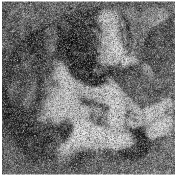
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

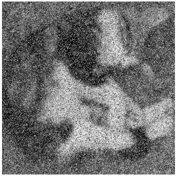
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

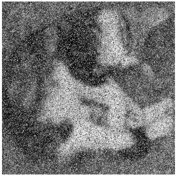
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

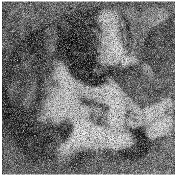
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

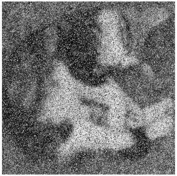
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

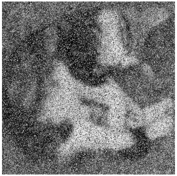
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

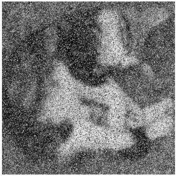
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

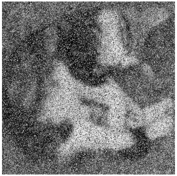
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

Iterative regularization at work

Recall that

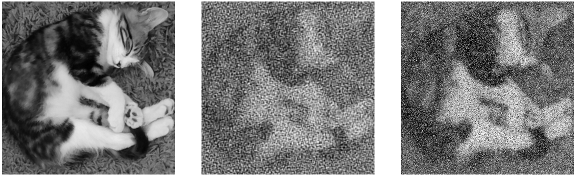
$$\|\widehat{x}_t - x^\dagger\| \leq \underbrace{\|\widehat{x}_t - x_t\|}_{\text{stability}} + \underbrace{\|x_t - x^\dagger\|}_{\text{optimization}}$$

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

Iterative regularization at work

Recall that

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



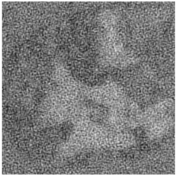
original image $\hat{x}_t$ $\hat{y}$

Iterative regularization at work

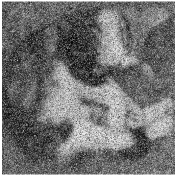
Recall that

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

original image



$\hat{x}_t$



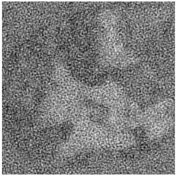
$\hat{y}$

Iterative regularization at work

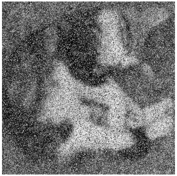
Recall that

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

original image



$\hat{x}_t$



$\hat{y}$

How to choose the stopping time?

How to choose the algorithm: duality

Let y be the exact datum.

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

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

How to choose the algorithm: duality

Let y be the exact datum.

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

where $\iota_{\{y\}}(x) = 0$ if $x = y$ and $\iota_{\{y\}}(x) = +\infty$ 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 the gradient method or an **inertial** gradient method to minimize it.

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

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

UniGe | MaLGa

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]

Theoretical guarantees & early stopping

Theorem (Matet, Rosasco, V., Vu, '17)

Let $D(Ax, y) = \|Ax - y\|^2$, and R be strongly convex. Let $\mathrm{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$ and $\|\widehat{y} - y\| \leq \delta$. If $\widehat{x}_t$ is the sequence generated by **gradient descent on the dual problem** associated to $\widehat{y}$, then

$$\|\widehat{x}_t - x^\dagger\| \leq c(1/\sqrt{t} + \sqrt{t}\delta).$$

Choosing $t_\delta \sim \delta^{-1}$, we have $\|\widehat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2})$

UniGe MaLGa

Theoretical guarantees & early stopping

Theorem (Matet, Rosasco, V., Vu, '17)

Let $D(Ax, y) = \|Ax - y\|^2$, and R be strongly convex. Let $\mathrm{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$ and $\|\widehat{y} - y\| \leq \delta$. If $\widehat{x}_t$ is the sequence generated by **gradient descent on the dual problem** associated to $\widehat{y}$, then

$$\|\widehat{x}_t - x^\dagger\| \leq c(1/\sqrt{t} + \sqrt{t}\delta).$$

Choosing $t_\delta \sim \delta^{-1}$, we have $\|\widehat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2})$

Remarks

- if $R = \|\cdot\|^2$ is Landweber algorithm

UniGe | MaLGa

Inertial version: Theoretical guarantees & early stopping

Theorem (Matet, Rosasco, V., Vu, '17)

Let $D(Ax, y) = \|Ax - y\|^2$, and R be strongly convex. Let $\operatorname{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$ and $\|\widehat{y} - y\| \leq \delta$. If $\widehat{x}_t$ is the sequence generated by **inertial gradient descent on the dual problem** associated to $\widehat{y}$, then

$$\|\widehat{x}_t - x^\dagger\| \leq c(1/t + t\delta).$$

Choosing $t_\delta \sim \delta^{-1/2}$, we have $\|\widehat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2})$

UniGe | MaLGa

Inertial version: Theoretical guarantees & early stopping

Theorem (Matet, Rosasco, V., Vu, '17)

Let $D(Ax, y) = \|Ax - y\|^2$, and R be strongly convex. Let $\mathrm{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$ and $\|\widehat{y} - y\| \leq \delta$. If $\widehat{x}_t$ is the sequence generated by **inertial gradient descent on the dual problem** associated to $\widehat{y}$, then

$$\|\widehat{x}_t - x^\dagger\| \leq c(1/t + t\delta).$$

Choosing $t_\delta \sim \delta^{-1/2}$, we have $\|\widehat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2})$

Remarks

- Same dependence on δ, but the good solution is reached faster
- if $R = \|\cdot\|^2$ see [Neubauer'16]

Inertial version: Theoretical guarantees & early stopping

Theorem (Matet, Rosasco, V., Vu, '17)

Let $D(Ax, y) = \|Ax - y\|^2$, and R be strongly convex. Let $\mathrm{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$ and $\|\widehat{y} - y\| \leq \delta$. If $\widehat{x}_t$ is the sequence generated by **inertial gradient descent on the dual problem** associated to $\widehat{y}$, then

$$\|\widehat{x}_t - x^\dagger\| \leq c(1/t + t\delta).$$

Choosing $t_\delta \sim \delta^{-1/2}$, we have $\|\widehat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2})$

Extension to general D?

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

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

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

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

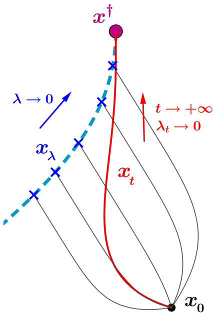
A diagonal approach[Lemaire 80s-90s]

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

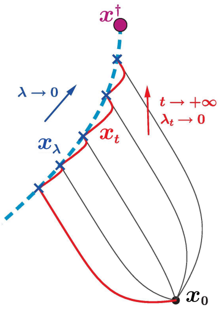
with $\lambda_t \to 0$.

A picture

The previous approach allows to describe:



A diagonal strategy



A warm restart strategy

A dual approach

Diagonal proximal gradient: [Attouch, Cabot, Czarnecki, Peypouquet ...]

A dual approach

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

A dual approach

Diagonal proximal gradient: [Attouch, Cabot, Czarnecki, Peypouquet ...]
Not well-suited if D is not smooth.
Assume $\operatorname{argmin} D(\cdot, y) = y$ and $D(y, y) = 0$.

A dual approach

Diagonal proximal gradient: [Attouch, Cabot, Czarnecki, Peypouquet ...]
Not well-suited if D is not smooth.
Assume $\operatorname{argmin} D(\cdot, y) = y$ and $D(y, y) = 0$.

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

Dual diagonal descent algorithm (3D)

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

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

We can use the **proximal gradient algorithm** on the dual.

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

UniGe | MaLGa

Dual diagonal descent algorithm (3D)

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

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

We can use the **proximal gradient algorithm** on the dual.

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

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

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

UniGe | MaLGa

Dual diagonal descent algorithm (3D)

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

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

We can use the **proximal gradient algorithm** on the dual.

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

UniGe | MaLGa

Convergence and stability of (3D)

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

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

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

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

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

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 $\operatorname{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$. Then

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

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

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 $\operatorname{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$. Then

- **Convergence** Let (x_t, u_t) be generated by (3D). Then:

$$\|x_t - x^\dagger\| \leq C/\sqrt{\sigma_R t}$$

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

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

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 $\operatorname{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$. Then

- **Convergence** Let (x_t, u_t) be generated by (3D). Then:

$$\|x_t - x^\dagger\| \leq C/\sqrt{\sigma_R t}$$

- **Stability** Suppose that $\|\hat{y} - y\| \leq \delta$. Let $\hat{x}_t$ be associated to $\hat{y}$. Then: $\|x_t - \hat{x}_t\| \leq C\delta t$.

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

AR) There exists $\bar{x}$ such that $A\bar{x} = y$ and $\bar{x} \in \mathrm{dom}R$.

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 $\mathrm{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$. Then

- **Convergence** Let (x_t, u_t) be generated by (3D). Then:

$$\|x_t - x^\dagger\| \leq C/\sqrt{\sigma_R t}$$

- **Stability** Suppose that $\|\hat{y} - y\| \leq \delta$. Let $\hat{x}_t$ be associated to $\hat{y}$. Then: $\|x_t - \hat{x}_t\| \leq C\delta t$.
- **Early stopping** There exists a stopping rule $t_\delta \sim \delta^{-2/3}$ such that

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

UniGe | MaLGa

Convergence and stability of (3D)

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

AD2) Let $p \in [1, +\infty]$. $D(\cdot, y)$ is p conditioned: $\frac{\gamma}{p}\|u - y\|^p \leq D(u, y)$

AR) There exists $\bar{x}$ such that $A\bar{x} = y$ and $\bar{x} \in \mathrm{dom} R$.

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 $\mathrm{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$. Then

- **Convergence** Let (x_t, u_t) be generated by (3D). Then:

$$\|x_t - x^\dagger\| \leq C/\sqrt{\sigma_R t}$$

- **Stability** Suppose that $\|\hat{y} - y\| \leq \delta$. Let $\hat{x}_t$ be associated to $\hat{y}$. Then: $\|x_t - \hat{x}_t\| \leq C\delta t$.
- **Early stopping** There exists a stopping rule $t_\delta \sim \delta^{-2/3}$ such that

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

Can be **accelerated**? Moreover, in general, not **optimal**

UniGe | MaLGa

Inertial dual diagonal descent algorithm (I3D)

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

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

We can use the **inertial proximal gradient algorithm** on the dual.

$$\begin{aligned}
& u_0 = u_1 \in G,\ \lambda_{\mathbf{t}} \to \mathbf{0}, \tau = \sigma_R/\|A\|^2,\ \alpha_t \nearrow 1 \\
& w_t = u_t + \alpha_t(u_t - u_{t-1}) \\
& z_{t+1} = w_t + \tau A \nabla R^*(-A^* w_t) \\
& u_{t+1} = \operatorname{prox}_{\tau\lambda_t^{-1}D^*(\lambda_t\cdot,y)}(z_{t+1}) = z_{t+1} - \tau \operatorname{prox}_{(\tau\lambda_t)^{-1}D(\cdot,y)}\left(\tau^{-1}z_{t+1}\right) \\
& x_{t+1} = \nabla R^*(-A^*u_t) = \operatorname{prox}_{\sigma_R^{-1}F}(-A^*u_t)
\end{aligned}$$

(I3D):convergence

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

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

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

(I3D):convergence

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

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

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

Theorem [Calatroni-Garrigos-Rosasco-V. 2019]

Suppose that $\lambda_t \in \ell^{1/(2(p-1))}(\mathbb{N})$. Let $x^\dagger$ be the solution of (P). Assume that $\operatorname{Im}(A^*) \cap \partial R(x^\dagger) \neq \varnothing$ and let x_t be generated by (I3D). Then:

$$\|x_t - x^\dagger\| \leq \frac{C}{\sqrt{\sigma_R}\, t}$$

(I3D):stability

Let $\hat{y}$ be such that $D(y, \hat{y}) \leq \delta^p$.

(I3D):stability

Let $\hat{y}$ be such that $D(y,\hat{y}) \leq \delta^p$.

$$\left|\begin{array}{l} \hat{u}_0 = \hat{u}_1 \in G,\ \ \lambda_{\mathbf{t}} \to \mathbf{0}, \tau = \sigma_R/\|A\|^2,\ \ \alpha_t \nearrow 1 \\ \hat{w}_t = \hat{u}_t + \alpha_t(\hat{u}_t - \hat{u}_{t-1}) \\ \hat{z}_{t+1} = \hat{w}_t + \tau A \nabla R^*(-A^* \hat{w}_t) \\ \hat{u}_{t+1} = \hat{z}_{t+1} - \tau \operatorname{prox}_{(\tau\lambda_t)^{-1} D(\cdot,\hat{y})}\left(\tau^{-1}\hat{z}_{t+1}\right) \\ \hat{x}_{t+1} = \operatorname{prox}_{\sigma_R^{-1} F}(-A^* \hat{u}_t) \end{array}\right.$$

UniGe | MaLGa

(I3D):stability

Let $\hat{y}$ be such that $D(y,\hat{y}) \leq \delta^p$.

$$
\left|\begin{array}{l}
\hat{u}_0 = \hat{u}_1 \in G,\ \lambda_{\mathbf{t}} \to \mathbf{0}, \tau = \sigma_R/\|A\|^2,\ \alpha_t \nearrow 1 \\
\hat{w}_t = \hat{u}_t + \alpha_t(\hat{u}_t - \hat{u}_{t-1}) \\
\hat{z}_{t+1} = \hat{w}_t + \tau A \nabla R^*(-A^* \hat{w}_t) \\
\hat{u}_{t+1} = \hat{z}_{t+1} - \tau\, \mathrm{prox}_{(\tau\lambda_t)^{-1} D(\cdot,\hat{y})} \left(\tau^{-1} \hat{z}_{t+1}\right) \\
\hat{x}_{t+1} = \mathrm{prox}_{\sigma_R^{-1} F}(-A^* \hat{u}_t)
\end{array}\right.
$$

To study stability we need to study **convergence under perturbations of the prox operator**

[Rockafellar '76; Le Roux-Schmidt-Bach 11; Salzo-V. '12; V.-Salzo-Baldassarre-Verri'13; Aujol-Dossal '15]

UniGe | MaLGa

Interlude: inexact prox

$$p = \operatorname{prox}_f(x) \iff p = \operatorname{argmin}\{f(z) + \frac{1}{2}\|x - z\|^2\} \iff x - p \in \partial f(p)$$

Possible notions of approximation [Salzo-V. '12]:

Interlude: inexact prox

$$p = \mathrm{prox}_f(x) \iff p = \mathrm{argmin}\{f(z) + \frac{1}{2}\|x - z\|^2\} \iff x - p \in \partial f(p)$$

Possible notions of approximation [Salzo-V. '12]:

Type 1 errors: $x + e - \hat{p} \in \partial_{\varepsilon_2^2/2} f(\hat{p})$, $\|e\| \leq \varepsilon_3$, and $\varepsilon_2^2 + \varepsilon_3^2 \leq \varepsilon$.

Interlude: inexact prox

$$p = \operatorname{prox}_f(x) \iff p = \operatorname{argmin}\{f(z) + \frac{1}{2}\|x - z\|^2\} \iff x - p \in \partial f(p)$$

Possible notions of approximation [Salzo-V. '12]:

Type 2 errors: $x - \hat{p} \in \partial_{\varepsilon^2/2} f(\hat{p})$

UniGe | MaLGa

Interlude: inexact prox

$$p = \mathrm{prox}_f(x) \iff p = \mathrm{argmin}\{f(z) + \frac{1}{2}\|x - z\|^2\} \iff x - p \in \partial f(p)$$

Possible notions of approximation [Salzo-V. '12]:

Type 3 errors: $x + e - \hat{p} \in \partial f(\hat{p}), \|e\| \leq \varepsilon.$

UniGe | MaLGa

Interlude: inexact prox

$$p = \mathrm{prox}_f(x) \iff p = \mathrm{argmin}\{f(z) + \frac{1}{2}\|x - z\|^2\} \iff x - p \in \partial f(p)$$

Possible notions of approximation [Salzo-V. '12]:

Type 1 errors: $x + e - \hat{p} \in \partial_{\varepsilon_2^2/2} f(\hat{p})$, $\|e\| \leq \varepsilon_3$, and $\varepsilon_2^2 + \varepsilon_3^2 \leq \varepsilon$.

Type 2 errors: $x - \hat{p} \in \partial_{\varepsilon^2/2} f(\hat{p})$

Type 3 errors: $x + e - \hat{p} \in \partial f(\hat{p}), \|e\| \leq \varepsilon$.

UniGe | MaLGa

Interlude: inexact prox

$$p = \mathrm{prox}_f(x) \iff p = \mathrm{argmin}\{f(z) + \frac{1}{2}\|x - z\|^2\} \iff x - p \in \partial f(p)$$

Possible notions of approximation [Salzo-V. '12]:

Type 1 errors: $x + e - \hat{p} \in \partial_{\varepsilon_2^2/2} f(\hat{p}), \quad \|e\| \leq \varepsilon_3$, and $\varepsilon_2^2 + \varepsilon_3^2 \leq \varepsilon$.

Type 2 errors: $x - \hat{p} \in \partial_{\varepsilon^2/2} f(\hat{p})$

Type 3 errors: $x + e - \hat{p} \in \partial f(\hat{p}), \|e\| \leq \varepsilon$.

Remark: If $\hat{p}$ is an approximation of type 2 or 3 then it is of type 1.

Proposition (Calatroni, Garrigos, Rosasco, V. '19)

- **Additive losses** correspond to **type 3** errors, with $\varepsilon = O(\delta)$.
- **KL** corresponds to **type 2** errors, with $\varepsilon = O(\delta/\lambda)$.

Error estimates

Lemma: Error estimates [Calatroni, Garrigos, Rosasco, V. - 2019]

Assume that $\lambda_t \in \ell^{\frac{1}{2(p-1)}}(\mathbb{N}))$. Let $\hat{x}_t$ be the I3D sequence associated to $\widehat{y}$ and assume that $\operatorname{prox}_{(\tau\lambda_t)^{-1}D(\cdot,\hat{y})}\left(\tau^{-1}\hat{z}_{t+1}\right)$ is a type 1 approximation. Then:

$$(\forall t \in \mathbb{N}) \quad \|\hat{x}_t - x^\dagger\|^2 \leq \frac{C}{\sigma_R t^2}\left\{1 + \left[\sum_{j=1}^{t-1} j^2 \varepsilon_{2,j}^2 + \frac{5}{2}\left(\sum_{j=1}^{k-1} j\varepsilon_{3,j}\right)^2\right]\right\}$$

- Joint **convergence** and **stability** estimate
- Very general (describes several perturbations)
- Idea of the proof: Lyapunov

UniGe | MaLGa

Iterative regularization results

Early stopping [Calatroni, Garrigos, Rosasco, V - 2019]

Assume that $\lambda_t \in \ell^{\frac{1}{2(p-1)}}(\mathbb{N})$. Let $\hat{x}_t$ be the I3D sequence associated to $\widehat{y}$, with $D(y, \hat{y}) \leq \delta^p$. Then, if :

- Let $D(z, y) = L(z - y)$. If $t_\delta \sim \delta^{-1/2}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2}).$$

UniGe | MaLGa

Iterative regularization results

Early stopping [Calatroni, Garrigos, Rosasco, V - 2019]

Assume that $\lambda_t \in \ell^{\frac{1}{2(p-1)}}(\mathbb{N})$. Let $\hat{x}_t$ be the I3D sequence associated to $\widehat{y}$, with $D(y, \hat{y}) \leq \delta^p$. Then, if :

- Let $D(z, y) = L(z - y)$. If $t_\delta \sim \delta^{-1/2}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2}).$$

- Let $D(z, y) = KL(y, z)$ (in this case $q = 2$) and $\lambda_t = t^{-\theta}$, with $\theta > 2$. If $t_\delta \sim \delta^{-2/(3+2\theta)}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{2/(3+2\theta)}).$$

Iterative regularization results

Early stopping [Calatroni, Garrigos, Rosasco, V - 2019]

Assume that $\lambda_t \in \ell^{\frac{1}{2(p-1)}}(\mathbb{N})$. Let $\hat{x}_t$ be the I3D sequence associated to $\widehat{y}$, with $D(y,\hat{y}) \leq \delta^p$. Then, if :

- Let $D(z,y) = L(z-y)$. If $t_\delta \sim \delta^{-1/2}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2}).$$

- Let $D(z,y) = KL(y,z)$ (in this case $q=2$) and $\lambda_t = t^{-\theta}$, with $\theta > 2$. If $t_\delta \sim \delta^{-2/(3+2\theta)}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{2/(3+2\theta)}).$$

Earl**ier** stopping

UniGe | MaLGa

Iterative regularization results

Early stopping [Calatroni, Garrigos, Rosasco, V - 2019]

Assume that $\lambda_t \in \ell^{\frac{1}{2(p-1)}}(\mathbb{N})$. Let $\hat{x}_t$ be the I3D sequence associated to $\widehat{y}$, with $D(y, \hat{y}) \leq \delta^p$. Then, if :

- Let $D(z, y) = L(z - y)$. If $t_\delta \sim \delta^{-1/2}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{1/2}).$$

- Let $D(z, y) = KL(y, z)$ (in this case $q = 2$) and $\lambda_t = t^{-\theta}$, with $\theta > 2$. If $t_\delta \sim \delta^{-2/(3+2\theta)}$ then

$$\|\hat{x}_{t_\delta} - x^\dagger\| = O(\delta^{2/(3+2\theta)}).$$

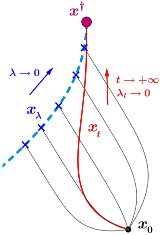
Earl**ier** stopping

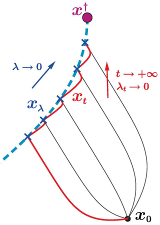
Optimal dependence on δ for quadratic losses

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 $= (\lambda_t) =$ n. iter.

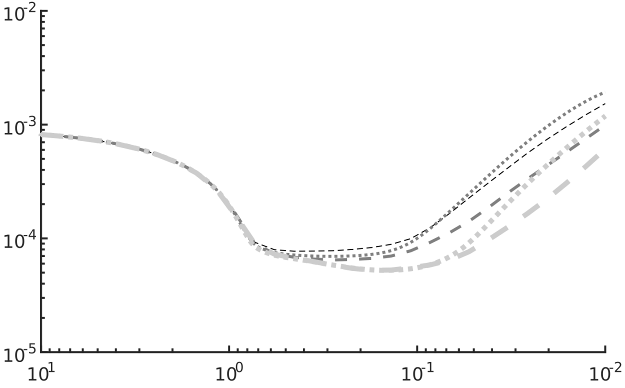
warm restart:

2 parameters: (λ_t); accuracy

UniGe | MaLGa

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

Euclidean distance from the true image

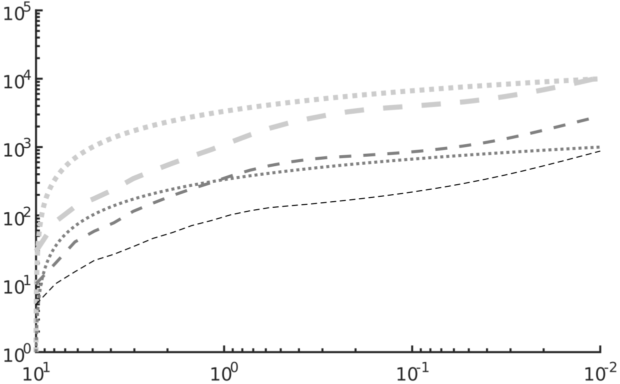


Dotted lines: diagonal with 10^3 and 10^4 iterations

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: $10^{-3}, 10^{-4}, 10^{-5}$

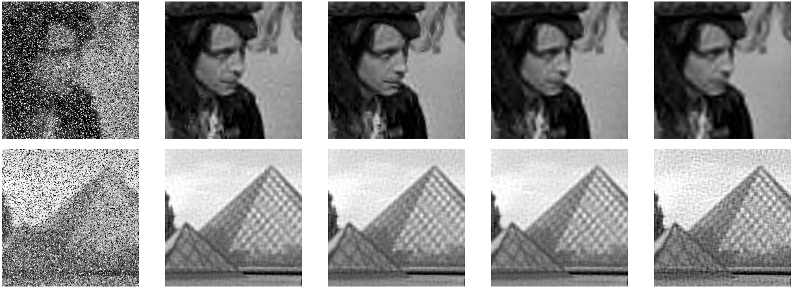
UniGe | MaLGa

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

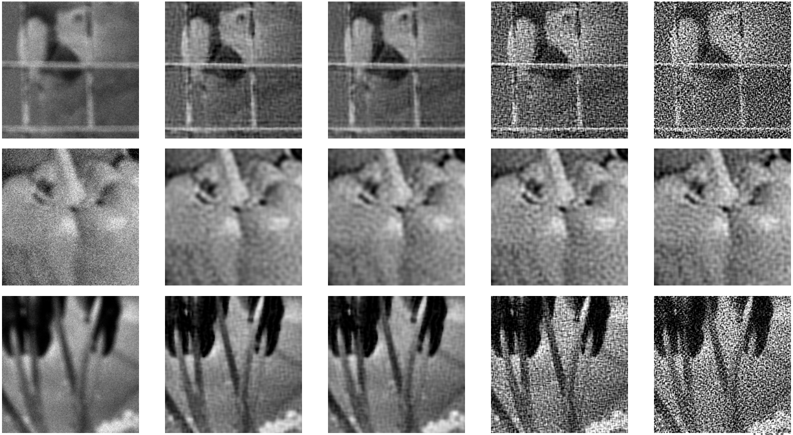
Blurring + Salt and pepper 35%. $D(u, y) = \|u - y\|_1$,
$R(x) = \|Wx\|_1 + \|x\|^2$ or $\|x\|_{TV} + \|x\|^2$



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

- remove strong convexity
- better use of conditioning?

References

- S. Matet, L. Rosasco, S. Villa, B. C. Vũ, Don't relax: early stopping for convex regularization, arxiv 2017.
- G. Garrigos, L. Rosasco, and S. Villa, Iterative regularization via dual diagonal descent, JMIV 2018
- L. Calatroni, G. Garrigos, L. Rosasco, and S. Villa, Iterative regularization via inertial dual diagonal descent, manuscript 2019

UniGe | MaLGa