Iterative regularization via dual diagonal descent

Silvia Villa

Department of Mathematics
University of Genoa

IHP, Paris, April 1st, 2019

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: **L. Calatroni, G. Garrigos, L. Rosasco**

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 $y \in G$. We have access only to $\widehat{y} \in G$ such that

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

Goal:

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

using only $\widehat{y}$.

Noisy data

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

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

Goal:

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

using only $\widehat{y}$.
Constraint $Ax = y$ can be replaced by

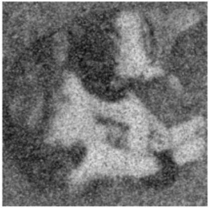
$$x \in \underset{x'}{\operatorname{argmin}}\, D(Ax', y)$$

for a **data fit** function D.

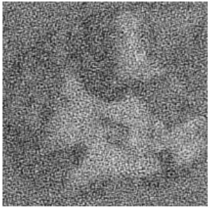
Regularization

original image x

$y = Ax$



$\hat{y}$

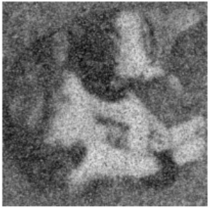


$\hat{x}^{\dagger}$

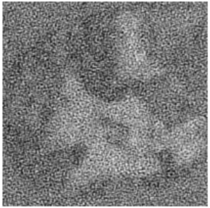
Regularization

original image x

$y = Ax$



$\hat{y}$



$\hat{x}^\dagger$

Regularization is needed

Tikhonov regularization

$$\underset{x \in \operatorname{argmin} D(A\cdot,\hat{y})}{\text{minimize}} \quad R(x) \qquad \rightarrow \qquad \underset{x \in \mathcal{H}}{\text{minimize}} \ \frac{1}{\lambda} D(Ax, \hat{y}) + R(x)$$

Tikhonov regularization

$$\underset{x \in \operatorname{argmin} D(A\cdot,\hat{y})}{\text{minimize}} \quad R(x) \qquad \rightarrow \qquad \underset{x \in \mathcal{H}}{\text{minimize}} \quad \frac{1}{\lambda} D(Ax, \hat{y}) + R(x)$$

How to choose λ?

Tikhonov regularization

$$\underset{x \in \operatorname{argmin} D(A\cdot,\hat{y})}{\text{minimize}} \quad R(x) \qquad \rightarrow \qquad \underset{x \in \mathcal{H}}{\text{minimize}} \ \ \frac{1}{\lambda} D(Ax, \hat{y}) + R(x)$$

How to choose λ?

Theorem

Let $D(Ax, y) = \|Ax - y\|^2$ and let $\widehat{x}^{\lambda}$ be the solution of the regularized problem. Suppose that there exists $q \in G$ such that $A^* q \in \partial R(x^\dagger)$. 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 $[\lambda_{\min}, \lambda_{\max}]$
- **optimization:** approximately solve the regularized problem for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ up to a certain accuracy ϵ
- **parameter selection:** select the best λ according to a validation criterion

Tikhonov regularization

What about computations?

Tikhonov regularization in practice:

- choose an interval $[\lambda_{\min}, \lambda_{\max}]$
- **optimization:** approximately solve the regularized problem for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ up to a certain accuracy ϵ
- **parameter selection:** select the best λ according to a validation criterion

Another point of view: integrate

REGULARIZATION and **OPTIMIZATION**

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

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

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

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

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

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

$$\widehat{x}_t \searrow$$

Iterative regularization: idea of the proof

Recall that

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

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

$$\widehat{x}_t \searrow \widehat{x}_{t+1} \searrow \dots \searrow$$

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

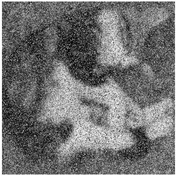
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

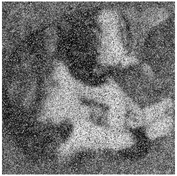
Iterative regularization at work

Recall that

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

original image

$\hat{x}_t$



$\hat{y}$

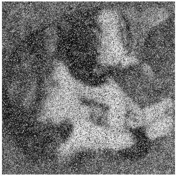
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

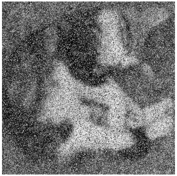
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

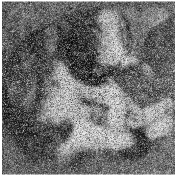
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

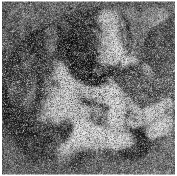
Iterative regularization at work

Recall that

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

original image

$\hat{x}_t$



$\hat{y}$

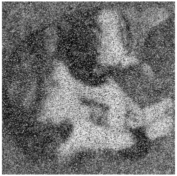
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

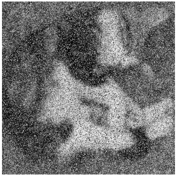
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

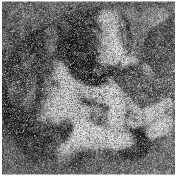
Iterative regularization at work

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

$\hat{x}_t$



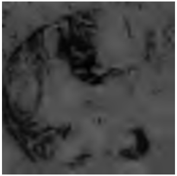
$\hat{y}$

Iterative regularization at work

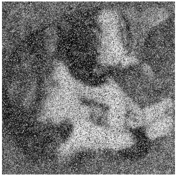
Recall that

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

original image



$\hat{x}_t$



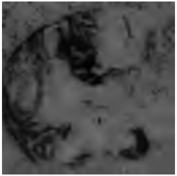
$\hat{y}$

Iterative regularization at work

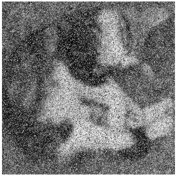
Recall that

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

original image



$\hat{x}_t$



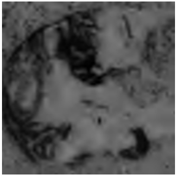
$\hat{y}$

Iterative regularization at work

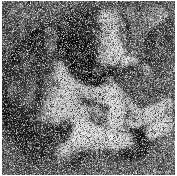
Recall that

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

original image



$\hat{x}_t$



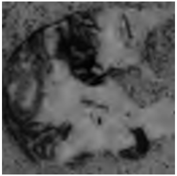
$\hat{y}$

Iterative regularization at work

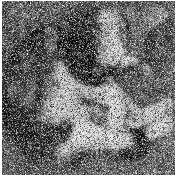
Recall that

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

original image



$\hat{x}_t$

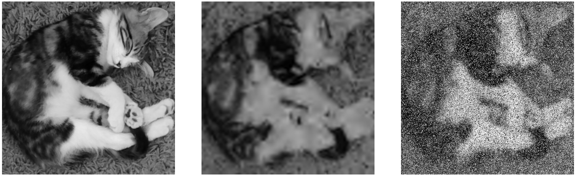


$\hat{y}$

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

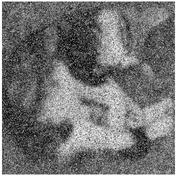
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

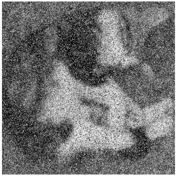
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

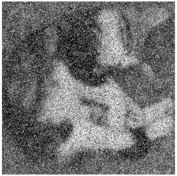
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

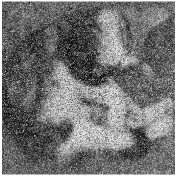
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

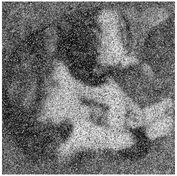
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

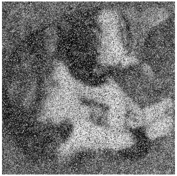
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

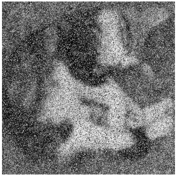
Iterative regularization at work

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

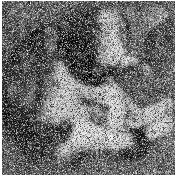
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

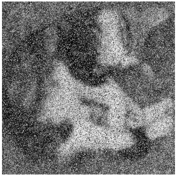
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

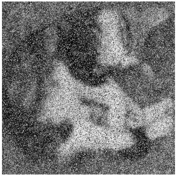
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

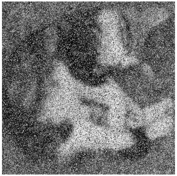
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

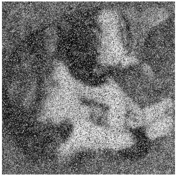
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

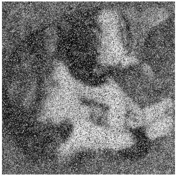
Iterative regularization at work

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


$\hat{x}_t$


$\hat{y}$

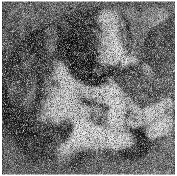
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

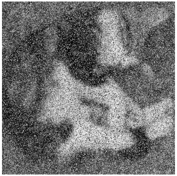
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

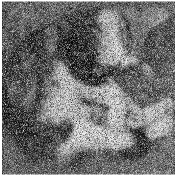
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

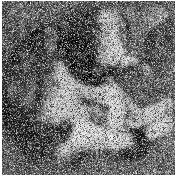
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

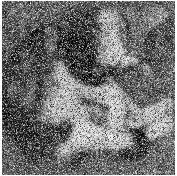
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

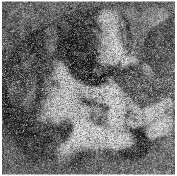
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

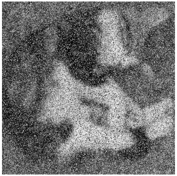
Iterative regularization at work

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

$\hat{x}_t$



$\hat{y}$

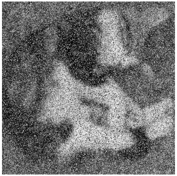
Iterative regularization at work

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

$\hat{x}_t$



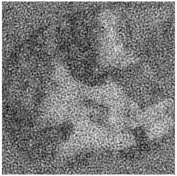
$\hat{y}$

Iterative regularization at work

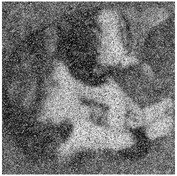
Recall that

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

original image



$\hat{x}_t$



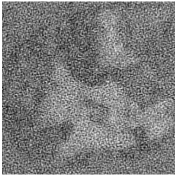
$\hat{y}$

Iterative regularization at work

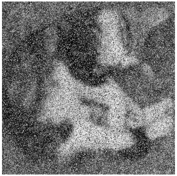
Recall that

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

original image



$\hat{x}_t$



$\hat{y}$

Dual problem

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

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

Dual problem

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

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

The dual problem is

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

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

Dual gradient descent

We can apply gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ u_{t+1} = u_t + \gamma(Ax_t - y) \end{cases}$$

Dual gradient descent

We can apply gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ u_{t+1} = u_t + \gamma(Ax_t - y) \end{cases}$$

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

Dual gradient descent

We can apply gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ u_{t+1} = u_t + \gamma(Ax_t - y) \end{cases}$$

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

When $R = \|\cdot\|^2/2$, it becomes the Landweber algorithm

$$x_{t+1} = (I - \gamma A^* A)x_t + \gamma A^* y$$

Dual gradient descent

We can apply gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ u_{t+1} = u_t + \gamma(Ax_t - y) \end{cases}$$

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

When $R = \|\cdot\|^2/2$, it becomes the Landweber algorithm

$$x_{t+1} = (I - \gamma A^* A)x_t + \gamma A^* y$$

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

Accelerated dual gradient descent

We can apply an accelerated gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ z_t = \nabla R^*\big(- A^* w_t\big) \\ w_t = u_t + \alpha_t (u_t - u_{t-1}) \\ u_{t+1} = w_t + \gamma (A z_t - y) \end{cases} , \quad \alpha_t = \frac{t-1}{t+\alpha}, \ \alpha \geq 2$$

Accelerated dual gradient descent

We can apply an accelerated gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ z_t = \nabla R^*\big(-A^* w_t\big) \\ w_t = u_t + \alpha_t(u_t - u_{t-1}) \\ u_{t+1} = w_t + \gamma(Az_t - y) \end{cases} \quad , \quad \alpha_t = \frac{t-1}{t+\alpha},\ \alpha \geq 2$$

When $R = \|\cdot\|^2/2$, it becomes an accelerated Landweber algorithm

$$\begin{cases} z_t = x_t + \alpha_t(x_t - x_{t-1}) \\ x_{t+1} = z_t - \gamma A^*(Az_t - y) \end{cases}$$

Accelerated dual gradient descent

We can apply an accelerated gradient descent to the dual problem:

$$\begin{cases} x_t = \nabla R^*(-A^* u_t) \\ z_t = \nabla R^*\big(-A^* w_t\big) \\ w_t = u_t + \alpha_t(u_t - u_{t-1}) \\ u_{t+1} = w_t + \gamma(Az_t - y) \end{cases} \quad , \quad \alpha_t = \frac{t-1}{t+\alpha}, \; \alpha \geq 2$$

When $R = \|\cdot\|^2/2$, it becomes an accelerated Landweber algorithm

$$\begin{cases} z_t = x_t + \alpha_t(x_t - x_{t-1}) \\ x_{t+1} = z_t - \gamma A^*(Az_t - y) \end{cases}$$

Accelerated gradient 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 $q \in G$ 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 $q \in G$ such that

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

Same condition needed for establishing rates for Tikhonov regularization.

Dual gradient descent is a regularization method

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

Assume that there $\exists q \in G$ *such that* $A^* q \in \partial R(x^\dagger)$. *Let* $u^\dagger$ *be a solution of the dual problem. For every* $\delta > 0$ *there exists* $t_\delta \sim \delta^{-1}$ *such that*

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

Accelerated dual gradient descent is a regularization method

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

Assume that there $\exists q \in G$ *such that* $A^*q \in \partial R(x^\dagger)$. *Let* $u^\dagger$ *be a solution of the dual problem. For every* $\delta > 0$ *there exists* $t_\delta \sim \delta^{-1/2}$ *such that*

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

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

Accelerated dual gradient descent is a regularization method

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

Assume that there $\exists q \in G$ such that $A^* q \in \partial R(x^\dagger)$. Let $u^\dagger$ be a solution of the dual problem. For every $\delta > 0$ there exists $t_\delta \sim \delta^{-1/2}$ such that

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

Based on the results of [Aujol-Dossal, 2016]
For $R = \|\cdot\|^2/2$ see also [A. Neubauer 2016]

Accelerated dual gradient descent is a regularization method

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

Assume that there $\exists q \in G$ such that $A^* q \in \partial R(x^\dagger)$. Let $u^\dagger$ be a solution of the dual problem. For every $\delta > 0$ there exists $t_\delta \sim \delta^{-1/2}$ such that

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

Based on the results of [Aujol-Dossal, 2016]
For $R = \|\cdot\|^2/2$ see also [A. Neubauer 2016]

What is the difference?

Accelerated dual gradient descent is a regularization method

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

Assume that there $\exists q \in G$ *such that* $A^* q \in \partial R(x^\dagger)$. *Let* $u^\dagger$ *be a solution of the dual problem. For every* $\delta > 0$ *there exists* $t_\delta \sim \delta^{-1/2}$ *such that*

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

Based on the results of [Aujol-Dossal, 2016]
For $R = \|\cdot\|^2/2$ see also [A. Neubauer 2016]

What is the difference?

Gradient descent: $t_\delta \sim \delta^{-1}$ Accelerated gradient descent: $t_\delta \sim \delta^{-1/2}$

General data fit

If $D(Ax, y) \neq \|Ax - y\|^2$ the previous approach does not work.
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

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

General data fit

If $D(Ax, y) \neq \|Ax - y\|^2$ the previous approach does not work.
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

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

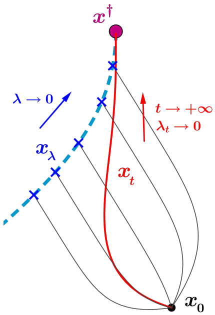
A diagonal approach[Lemaire 80s-90s]

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

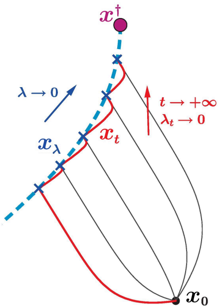
with $\lambda_t \to 0$.

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
- Require to know the conditioning of $D(A\cdot; y)$ (might not exists)

A dual approach

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

- Not well-suited if D is not smooth
- Require to know the conditioning of $D(A\cdot;y)$ (might not exists)

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

Dual diagonal descent algorithm (3D)

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

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

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

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

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.

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

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

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

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.

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

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 $A\bar{x} = y$ $\quad \bar{x} \in \operatorname{dom} R$.

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 $A\bar{x} = y$ $\quad\bar{x} \in \operatorname{dom} R$.

Theorem [Garrigos-Rosasco-V. 2017]

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

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

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

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

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

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

Accelerated dual diagonal descent algorithm (A3D)

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 **accelerated forward-backward splitting algorithm** on the dual.

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

(A3D) is a regularization method

Theorem (Early-stopping) [Calatroni-Garrigos-Rosasco-V. 2019]

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 (A3D) algorithm with $y = \hat{y}$ and $\hat{u}_0 = u_0$. Suppose that

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

Then there exists an early stopping rule $t(\delta) \sim \delta^{-1/2}$ which verifies

$$\|\hat{x}_{t(\delta)} - x^\dagger\| = O(\delta^{\frac{1}{2}}) \text{ when } \delta \to 0.$$

(A3D) is a regularization method

Theorem (Early-stopping) [Calatroni-Garrigos-Rosasco-V. 2019]

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 (A3D) algorithm with $y = \hat{y}$ and $\hat{u}_0 = u_0$. Suppose that

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

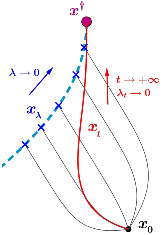
Then there exists an early stopping rule $t(\delta) \sim \delta^{-1/2}$ which verifies

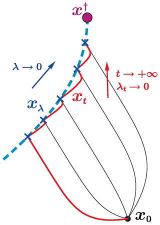
$$\|\hat{x}_{t(\delta)} - x^\dagger\| = O(\delta^{\frac{1}{2}}) \text{ when } \delta \to 0.$$

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

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**



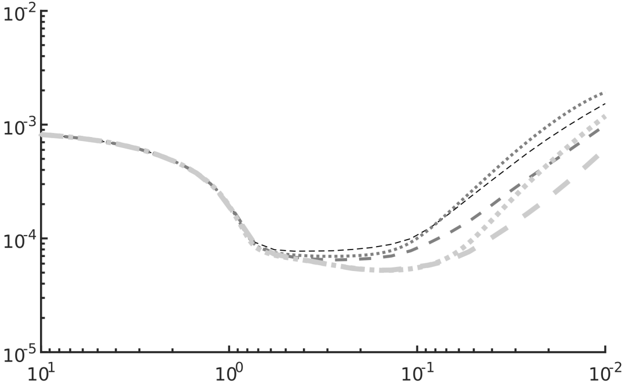


diagonal:
one parameter = (λ_t)= n. iter.

warm restart:
two parameters: (λ_t); accuracy

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

Euclidean distance from the true image

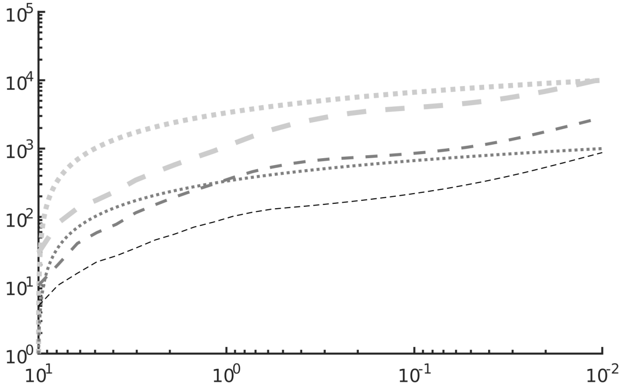


Dotted lines: diagonal with 10^3 and 10^4 iterations

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

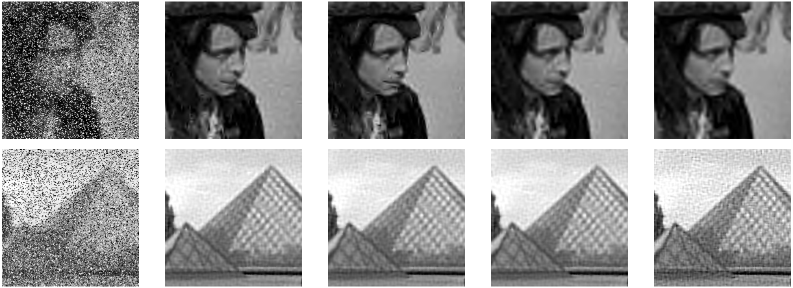
Dashed lines: warm restart with 30 λs and accuracy: $10^{-3}, 10^{-4}, 10^{-5}$

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

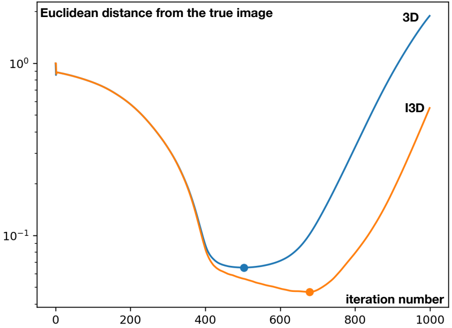
Blurring + Salt and pepper 35%. $D(u,y) = \|u - y\|_1$,
$R(x) = \|Wx\|_1 + \|x\|^2$ or $\|x\|_{TV} + \|x\|^2$

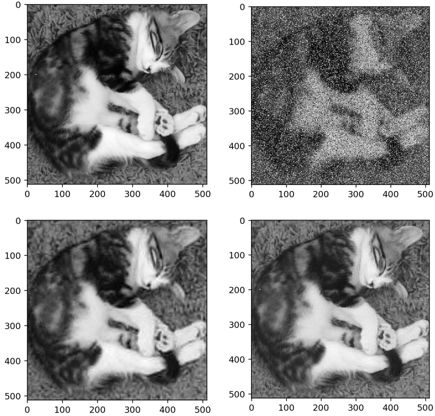


noisy image, reconstruction with diagonal and warm restart using true image, reconstruction with diagonal and warm restart using SURE

Comparison between 3D and A3D

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





3D vs. A3D


original image


$3D$


$A3D$

350
Iterations

3D vs. A3D

original image

$3D$

$A3D$

400
Iterations

3D vs. A3D

original image

$3D$

$A3D$

430
Iterations

3D vs. A3D

original image

$3D$

$A3D$

500
Iterations

3D vs. A3D

original image

$3D$

$A3D$

680
Iterations

3D vs. A3D

original image

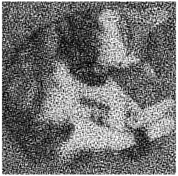
$3D$

$A3D$

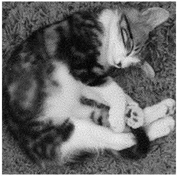
750
Iterations

3D vs. A3D

original image



$3D$



$A3D$

900
Iterations

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
- learning setting $(A \to \widehat{A})$?

References

S. Matet, L. Rosasco, 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, Accelerated iterative regularization via dual diagonal descent, manuscript 2019

The end

Merci pour votre attention