Regularization of inverse problems in image processing

Khalid Jalalzai

CMAP, École Polytechnique

March 9, 2012

Introduction

Inverse problems in imaging

A damaged image $g : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}$ is represented as:

$$g = Ag_0 + n.$$

Our aim: restore the image!

Restoring by minimizing an energy

Various approaches: Partial Differential Equations, Statistical estimators, Sparse representations, Variational methods.

Restoring by minimizing an energy

Various approaches: Partial Differential Equations, Statistical estimators, Sparse representations, **Variational methods.**

Often, one minimizes an energy of the form

$$\mathcal{E}(u) = \frac{1}{2}\|Au - g\|_2^2 + \lambda \mathcal{R}(u).$$

The first term behaves as a **data fidelity**, whereas $\mathcal{R}(u)$ is a **regularization** term that reflects an *a priori* distribution on images.

Penalizing oscillations

The idea: highly oscillating images are less probable.

In 1963, Tychonov suggested to minimize the following

$$\min_{u \in H^1(\Omega)} \frac{1}{2}\|Au - g\|_2^2 + \frac{\lambda}{2}\int_\Omega |\nabla u|^2.$$

In 1992, Rudin, Osher & Fatemi proposed the model

$$\min_{u \in BV(\Omega)} \frac{1}{2}\|Au - g\|_2^2 + \lambda TV(u), \qquad \text{(ROF)}$$

where $TV(u) = \int_\Omega |Du|$.

Penalizing oscillations

The idea: highly oscillating images are less probable.

In 1963, Tychonov suggested to minimize the following

$$\min_{u \in H^1(\Omega)} \frac{1}{2}\|Au - g\|_2^2 + \frac{\lambda}{2}\int_\Omega |\nabla u|^2.$$

In 1992, Rudin, Osher & Fatemi proposed the model

$$\min_{u \in BV(\Omega)} \frac{1}{2}\|Au - g\|_2^2 + \lambda TV(u), \qquad \text{(ROF)}$$

where $TV(u) = \int_\Omega |Du|$.

TV minimization

$\lambda = 10$

$\lambda = 30$

$\lambda = 100$

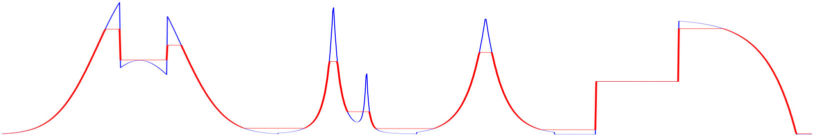
Fine Properties of the Total Variation Minimization Problem

ROF's model

For simplicity we consider the denoising problem

$$\min_{u \in BV(\Omega)} \frac{1}{2}\|u-g\|_2^2 + \lambda \int_\Omega |Du|.$$

- The TV term regularizes images without smoothing the edges of the objects.
- TV produces an undesirable artifact: the staircasing phenomenon.



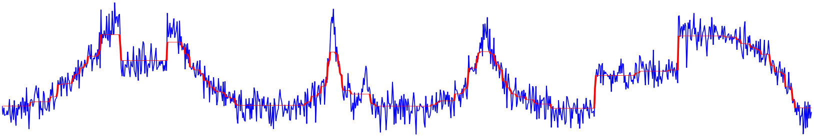
We are going to explore these properties further.

ROF's model

For simplicity we consider the denoising problem

$$\min_{u \in BV(\Omega)} \frac{1}{2}\|u - g\|_2^2 + \lambda \int_\Omega |Du|.$$

- The TV term regularizes images without smoothing the edges of the objects.
- TV produces an undesirable artifact: the staircasing phenomenon.



We are going to explore these properties further.

Discontinuities

Recently Caselles, Chambolle & Novaga (2008) showed that whenever $g \in L^\infty(\Omega) \cap BV(\Omega)$, the discontinuity set satisfy

$$J_u \subset J_g.$$

In a sense, **no new objects are created**.

An Anisotropic Energy

We generalized this result to energies of the form:

$$\mathcal{E}(u) = \int_\Omega \Phi(x, Du(x))dx + \int_\Omega \Psi(x, u(x))dx$$

where essentially

- Φ C^2 out of $\Omega \times \mathbb{R}^N \setminus \{0\}$, positively 1-homogenenous and elliptic in the second variable,
- Ψ measurable in the first variable, strictly convex and coercive in the second one.

Theorem

Assuming that for a countable set D dense in $\mathbb{R}$,

$$\partial_t \Psi(\cdot, t) \in BV(\Omega) \cap L^\infty(\Omega), \ \forall t \in D,$$

one has

$$J_u \subset \bigcup_{t \in D} J_{\partial_t \Psi(\cdot,t)}$$

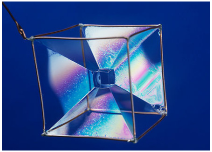
up to a set $\mathcal{H}^{N-1}$ negligible.

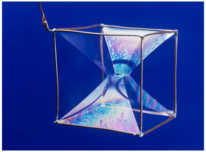
One can adapt the proof of CCN provided:

- One can understand how our problem relates to a **minimal surface problem**: denoting $E_s := \{u > s\}$

$$\int_\Omega \Phi(x, Du(x))dx + \int_\Omega \Psi(x, u(x))dx \sim \int_s \left(P_\Phi(E_s, \Omega) + \int_{E_s} \partial_t \Psi(x, t)dx \right) ds.$$

Two minimal surfaces:





- One can get the desired **regularity for the level sets** combining
  - the theory of regularity for quasi-minimal surfaces,
  - the Nirenberg's method,
  - the regularity theory for elliptic PDEs in non-divergence form.

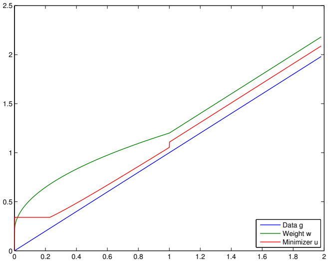
Refinement in the weighted case

Problem: What if the anisotropy is less regular?

For instance

$$\min_{u \in BV(\Omega)} \int_\Omega w|Du| + \frac{1}{2}\|u-g\|_2^2$$

with w merely Lipschitz continuous.



Creation of jumps with $w(x) = \sqrt{x}\chi_{\{x \leq 1\}} + x\chi_{\{x>1\}} + 0.2$

Theorem

Let $w : \Omega \to \mathbb{R}$ be positive, bounded, Lipschitz continuous with $\nabla w \in BV(\Omega, \mathbb{R}^N)$ and $g \in BV(\Omega) \cap L^\infty(\Omega)$.

Then the minimizer $u \in BV(\Omega)$ satisfies

$$J_u \subset J_g \cup J_{\nabla w}$$

up to a $\mathcal{H}^{N-1}$-negligible set.

If in addition we assume that w is of class C^1 we get that at the discontinuity

$$(u^+ - u^-) \leq (g^+ - g^-) \ \mathcal{H}^{N-1}\text{-a.e. on } J_u.$$

Theorem

Let $w : \Omega \to \mathbb{R}$ be positive, bounded, Lipschitz continuous with $\nabla w \in BV(\Omega, \mathbb{R}^N)$ and $g \in BV(\Omega) \cap L^\infty(\Omega)$.

Then the minimizer $u \in BV(\Omega)$ satisfies

$$J_u \subset J_g \cup J_{\nabla w}$$

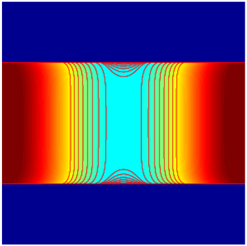
up to a $\mathcal{H}^{N-1}$-negligible set.

If in addition we assume that w is of class C^1 we get that at the discontinuity

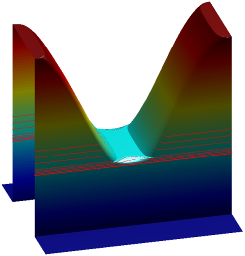
$$(u^+ - u^-) \leq (g^+ - g^-) \ \mathcal{H}^{N-1}\text{-a.e. on } J_u.$$

This is quite surprising if one thinks of

$$g : [0, 2\pi)^2 \longrightarrow \mathbb{R}$$
$$(x, y) \longmapsto \begin{cases} 2 + \cos(x) \text{ if } y > 0, \\ 0 \text{ otherwise.} \end{cases}$$



Level lines $\{u = t\}$ for some values of $t \in (1, 2)$.



Graph of u on one period.
Some level lines are represented in red.

Staircasing and discontinuities depend on λ

Not much can be said in general.

In 1D, Ring (2000) and Briani, Chambolle, Novaga, Orlandi (2011) show that the solutions $u(t)$ of

$$\min_{u \in BV(\Omega)} t \int_\Omega |Du| + \frac{1}{2}\|u-g\|_2^2.$$

form a semi-group.

Theorem

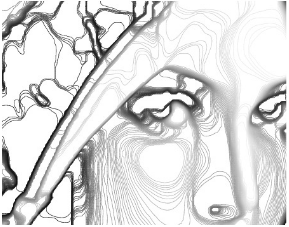
Let $\Omega = B(0,R) \subset \mathbb{R}^N$, $g \in L^2(\Omega)$ radial. Then $(u(t))_t$ form a semi-group.

Corollary

If $\lambda \leq \mu$, $J_\mu \subset J_\lambda$ and $S_\lambda \subset S_\mu$.

Discontinuities vanish, staircasing increases.

Staircasing



Level lines of a TV-minimizer

By looking at the level sets we prove that staircasing occurs

- at global extrema of g.
- at all extrema of u.

Some perspectives

Our work paved the way for future researches:

- Staircasing occurs a.e. for a noisy image.
- For a general g, do we have $J_\mu \subset J_\lambda$?
- Study the regularity of the minimizers in the anisotropic setting.

An Alternative for the Total Variation

A variant of TV

The idea: replace TV by

$$J(u,\Omega) = \inf_{P\varphi = Du} \int_\Omega |\varphi|,$$

where P is the "projection on gradients".

Remark that

$$J(u) \leq TV(u).$$

Motivates the use of J in image processing.

Dual formulation

Using Riesz's duality and some **convex analysis**:

Proposition

If $\Omega \subset \mathbb{R}^N$ is a convex open set then for any $u \in BV(\Omega)$,

$$J(u,\Omega) = \sup_{\substack{w \in C_c^1(\Omega) \\ \|\nabla w\|_\infty \leq 1}} \int_\Omega \nabla w \cdot Du.$$

Second order approach to reduce staircasing.

Theorem

Let $\Omega \subset \mathbb{R}^N$ open and $u = \chi_E$ the characteristic function of a set of finite perimeter E in Ω, or more generally $u \in BV(\Omega)$ with Du concentrated on the jump set J_u. Then,

$$J(u, \Omega) = \int_\Omega |Du|.$$

J coincides with TV on "cartoon" images.

The idea:

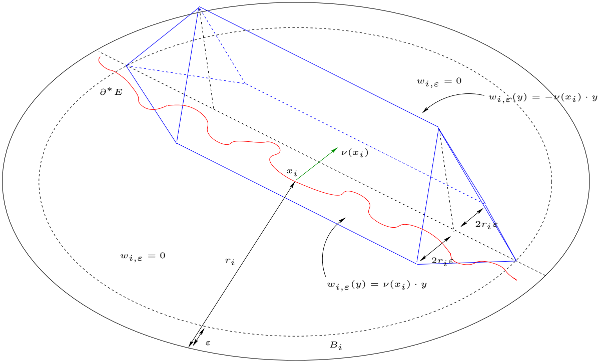
- If $u = \chi_E$ with ∂E a $C^{1,1}$ manifold.
  Consider the **signed distance** $w = d(x, \Omega \setminus E) - d(x, E)$.
  A classical result asserts that:

  $$w \text{ is } C^{1,1} \text{ near } \operatorname{supp}(Du) = \partial E \text{ and } \nabla w = \nu.$$

  Thus,

  $$J(u) \geq \int_\Omega \nabla w \cdot Du = \int_{\partial E} \nu \cdot Du = \int_\Omega |Du|.$$

- In the general case, we use some tools of **geometric measure theory** to:
  - localize the problem,
  - build w from scratch using the **rectifiability** of J_u.



ROF revisited

Given Ω open and $g \in L^2(\Omega)$, consider the problem

$$\min_{u \in L^2(\Omega)} \mathcal{F}(u) = \frac{1}{2}\|u - g\|_2^2 + \lambda J(u).$$

Proposition

$\mathcal{F}$ has a unique minimizer $u_\lambda \in L^2(\Omega)$.

Proposition (An explicit solution)

Let $g = C\chi_{B(0,1)}$ and $\lambda \geq 0$. Then, if $C \geq \lambda N$, the minimizer of $\mathcal{F}$ is

$$u_\lambda = (C - \lambda N)\chi_{B(0,1)}.$$

Numerical simulations: a noisy image

$\sigma = 20$

PSNR=22.1

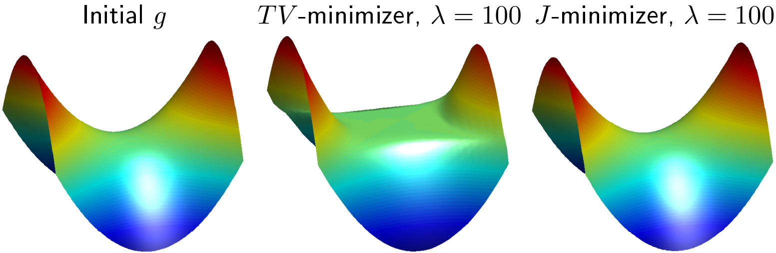
TV-minimizer, $\lambda = 25$

PSNR=29.4

$\tilde{J}$-minimizer, $\lambda = 25$

PSNR=29.3

Numerical simulations: absence of staircasing



Motivations and perspectives

- J behaves mostly like TV without creating homogeneous regions.
- Some open issues: Poincaré inequality, canonical space?

Adapted Basis for Non-Local Reconstruction of Spectrum

Non-Locality in images

Images have non-local features:

Non-Locality in images

Recently developed models take into account this structure:

- Denoising proposed by Buades, Coll, Morel (2005):

$$NLMeans(g)(x) = \frac{1}{C(x)} \int_\Omega g(y)w(x,y)dy$$

- Other inverse problems:

$$\min_u \frac{1}{2}\|Au - g\|_2^2 + \lambda \int_{\Omega\times\Omega} \|p_u(x) - p_u(y)\| w(x,y)dxdy$$

A key step is the computation of the similarity measure:

$$w(x,y) = \exp\left(-\frac{\|p_g(x) - p_g(y)\|_2}{h}\right).$$

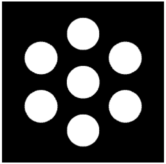
Spectrum reconstruction

The problem:

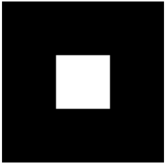
$$g = \mathcal{F}^{-1}(\chi_M \mathcal{F}(g_0))$$

Different masks M for various applications:

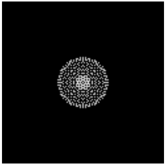
Spatial imaging



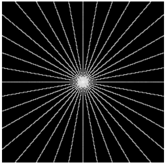
Zoom



Inverse Acoustic Scattering



Tomography



The aim: restore the spectrum.

NLMeans Similarity Measure

In general, $\delta(x,y) = \|p_g(x) - p_g(y)\|_2$:



Can we do better?

The aim: design a similarity measure $\delta(x,y)$ that is adapted to the problem of spectrum reconstruction.

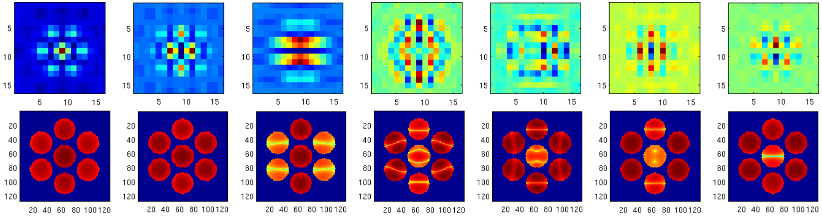
Adapted Atoms

The idea: design test functions $(\phi_\alpha)_\alpha$ such that

$$g * \phi_\alpha = g_0 * \phi_\alpha, \ \forall \alpha.$$

One can compute an orthogonal basis iteratively

$$\phi_\alpha = \text{argmin} \left\{ \int_\Omega |\phi(x)|^2 |x|_2^p dx, \ \text{supp}(\mathcal{F}\phi) \subset M, \|\phi\|_2 \geq 1, \phi \perp \text{Span}\{\phi_{\alpha'}, \alpha' < \alpha\} \right\}$$



Similarity measure comparison

We define the following similarity measure:

$$\delta(x,y) = \left(\sum_{\alpha \leq \alpha_0} |g * \phi_\alpha(x) - g * \phi_\alpha(y)|^2 \right)^{\frac{1}{2}}.$$

Here α_0 sets how localized the considered atoms are.

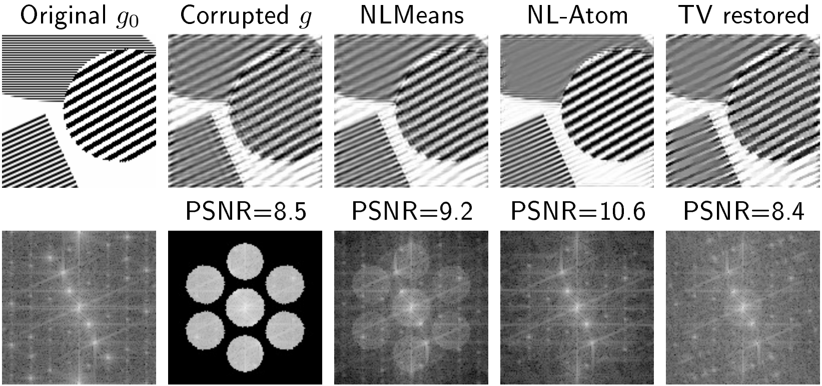
Performance of this new similarity measure

Atom distance

NLM distance

The 13 best matches (in red) for a fixed patch (in green).

Numerical simulations: a toy example



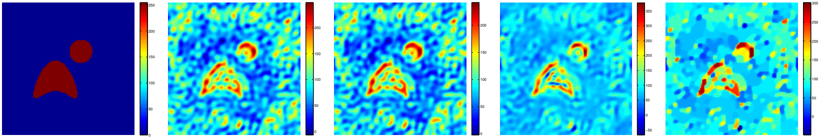
Numerical simulations: acoustic inverse scattering

Thanks to the Born approximation

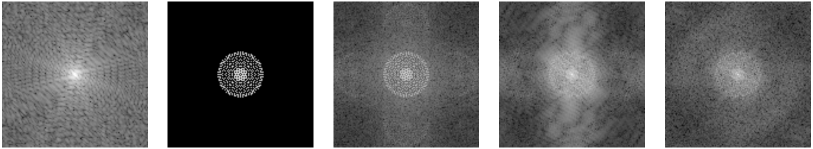
$$u_\infty(\hat{x}, d) \approx \int_{\mathbb{R}^N} \chi_D(y) e^{-ik(\hat{x}-d)\cdot y} dy,$$

we can use the data that comes out of the direct problem.
In a sense, we add noise.

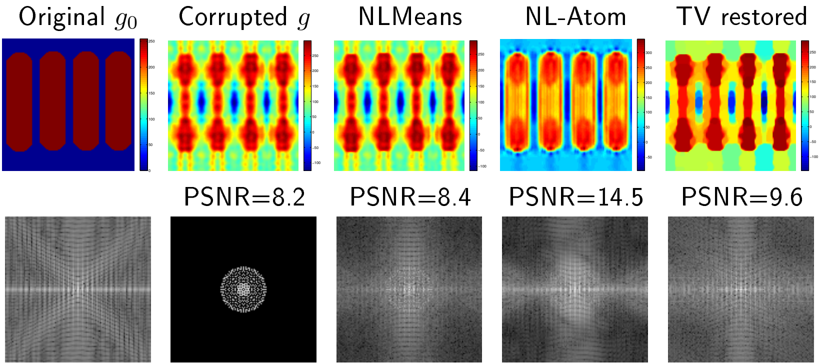
Original g_0 | Corrupted g | NLMeans | NL-Atom | TV restored



PSNR=9.24 | PSNR=9.25 | PSNR=9.81 | PSNR=9.42

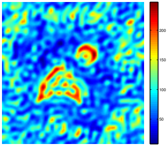


Numerical simulations: closely located objects

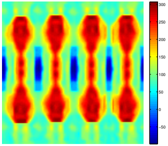


Numerical simulations: Weight recomputation


PSNR=12.1



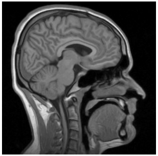
PSNR=9.27



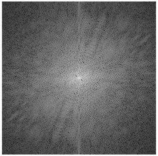
PSNR=8.9

Numerical simulations: Tomography problem

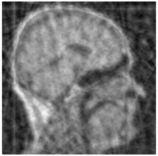
Original g_0



Spectrum of g_0

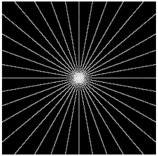


Corrupted g

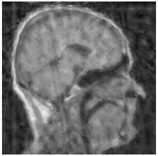


PSNR=22.4

Spectrum of g

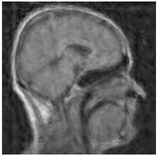


NLMeans



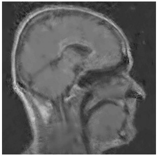
PSNR= 23.8

NL-Atom



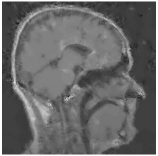
PSNR=24.9

NL-Atom then $1\times$ NLMeans



PSNR=25.8

$20\times$ NLMeans



PSNR=24.8

TV restored

PSNR=23.6

Advantages

- Performs much better in some cases.
- The weight computation is faster.
- The weight recomputation is not mandatory.

Convex Optimization: The Primal-Dual framework

Non-smooth minimization

Usually minimization is carried out by using gradient algorithms.

As far as we are concerned, we are interested in the **minimization of non-smooth energies** of the form

$$\min_{x \in X} F(Ax) + G(x).$$

- F lsc convex.
- G lsc uniformly convex with parameter γ_0.

New algorithms should be designed for such problems.

The Primal-Dual framework

The idea: consider a dual variable y.

A recently developed algorithm aims to find a saddle point $(\hat{x}, \hat{y})$ of the problem

$$\min_{x \in X} \max_{y \in Y} \langle Ax, y \rangle + G(x) - F^*(y)$$

and is inspired by the following

Algorithm 1 Arrow-Hurwicz's scheme

- **Iterations:** For $n \geq 1$ update as follows:

$$x^{n+1} = (I + \tau \partial G)^{-1}(x^n - \tau A^* y^n),$$
$$y^{n+1} = (I + \sigma \partial F^*)^{-1}(y^n + \sigma A x^{n+1}).$$

Adaptive stepsize

Chambolle, Pock (2010) propose the following modification:

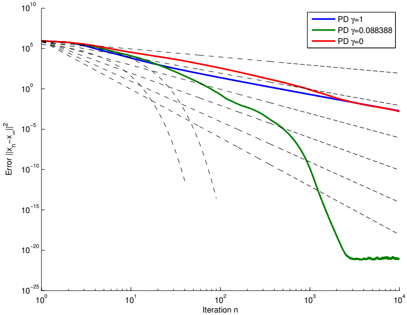
Algorithm 2 Primal Dual with adaptive stepsize

- **Initialization:** $\sigma_0\tau_0\|A\|^2 \leq 1$, $\gamma \leq \gamma_0$.
- **Iterations:** For $n \geq 1$, consider the updates:

$$
\begin{aligned}
y^{n+1} &= (I + \sigma_n \partial F^*)^{-1}(y^n + \sigma_n A\bar{x}^n), \\
x^{n+1} &= (I + \tau_n \partial G)^{-1}(x^n - \tau_n A^* y^{n+1}), \\
\theta_n &= 1/\sqrt{1 + 2\gamma\tau_n},\ \tau_{n+1} = \theta_n \tau_n,\ \sigma_{n+1} = \sigma_n/\theta_n, \\
\bar{x}^{n+1} &= x^{n+1} + \theta_n (x^{n+1} - x^n).
\end{aligned}
$$

Converges as $O\left(\frac{1}{n^2}\right)$.

Surprisingly the complexity depends on γ:



Error $\|x^n - \hat{x}\|^2$ for different values of γ

A first explanation

Algorithm 3 Primal Dual with adaptive stepsize

- **Initialization:** $\sigma_0\tau_0\|A\|^2 \leq 1$, $\gamma \leq \gamma_0$.
- **Iterations:** For $n \geq 1$, consider the following updates:

$$
\begin{aligned}
y^{n+1} &= (I + \sigma_n \partial F^*)^{-1}(y^n + \sigma_n A\bar{x}^n),\\
x^{n+1} &= (I + \tau_n \partial G)^{-1}(x^n - \tau_n A^* y^{n+1}),\\
\theta_n &= 1/\sqrt{1 + 2\gamma\tau_n},\ \tau_{n+1} = \theta_n \tau_n,\ \sigma_{n+1} = \sigma_n/\theta_n,\\
\bar{x}^{n+1} &= x^{n+1} + \theta_n(x^{n+1} - x^n).
\end{aligned}
$$

A first explanation

Algorithm 3 Primal Dual with adaptive stepsize

- **Initialization:** $\sigma_0\tau_0\|A\|^2 \leq 1$, $\gamma \leq \gamma_0$.
- **Iterations:** For $n \geq 1$, consider the following updates:

$$
\begin{aligned}
y^{n+1} &= (I+\sigma_n\partial F^*)^{-1}(y^n+\sigma_n A\bar{x}^n),\\
x^{n+1} &= (I+\tau_n\partial G)^{-1}(x^n-\tau_n A^* y^{n+1}),\\
\theta_n &= 1/\sqrt{1+\gamma\tau_n},\ \tau_{n+1}=\theta_n\tau_n,\ \sigma_{n+1}=\sigma_n/\theta_n,\\
\bar{x}^{n+1} &= x^{n+1}+\theta_n(x^{n+1}-x^n).
\end{aligned}
$$

One proved

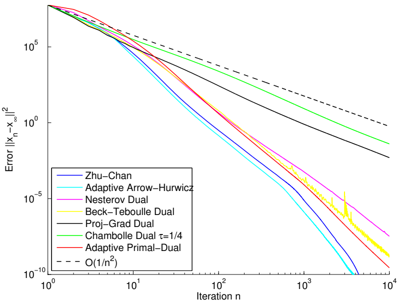
Theorem

Let $\tau_0, \sigma_0 > 0$ *such that* $\sigma_0 \tau_0 \|A\|^2 \leq 1$ *then the sequence* $(x^n)_{n\in\mathbb{N}}$ *converges to* $\hat{x}$ *and*

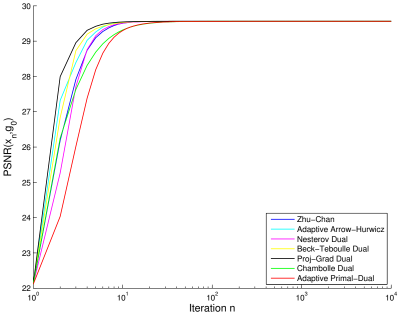
$$\sum_n n\|\hat{x} - x^n\|^2 < +\infty.$$

Complexity beyond $O\left(\frac{1}{n^2}\right)$: **best theoretical rate of convergence** for this class of problems.

Comparison for ROF's denoising problem



Minimizer error $\|x^n - \hat{x}\|^2$



PSNR(x^n, g_0)

Some perspectives

- Prove that the dual variable converges for the adaptive Primal-Dual algorithm.
- Devise the optimal uniform convexity parameter γ that gives the best rate and prove that it is beyond $o\left(\frac{1}{n^2}\right)$.

Merci !