Learning Nonexpansive Operators

Mathematical Models for Plug-and-Play Image Restoration
Paris, Centre Culturel Irlandais

Jonathan Chirinos Rodriguez
MaLGa- Machine learning Genova Center
Università di Genova
Camelot Biomedical Systems

Joint work with Prof. Kristian Bredies (University of Graz) and Emanuele Naldi (TU Braunschweig)

December, 7^{th}, 2022

The denoising problem

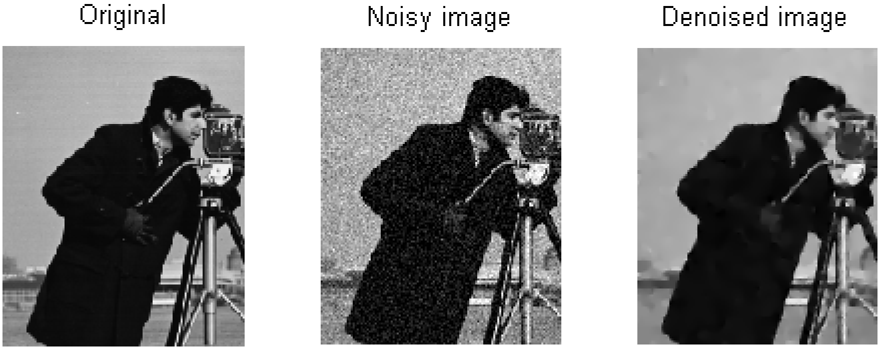


Figure: Denoising process with the ROF model[1]. Source: Wikipedia

[1] Rudin, L. I. ; Osher, S.; Fatemi, E., 1992.

Inverse problems

Let $\mathcal{X}$ be a separable Hilbert space. The denoising model writes

$$x = y^* + \varepsilon,$$

where x is the data, y^* the ground truth, and ε the noise.

Bauschke, H. H., Combettes, P. L., 2017.

Inverse problems

Let $\mathcal{X}$ be a separable Hilbert space. The denoising model writes

$$x = y^* + \varepsilon,$$

where x is the data, y^* the ground truth, and ε the noise.

Goal: Reconstruct y^* given x.

Bauschke, H. H., Combettes, P. L., 2017.

Inverse problems

Let $\mathcal{X}$ be a separable Hilbert space. The denoising model writes

$$x = y^* + \varepsilon,$$

where x is the data, y^* the ground truth, and ε the noise.

Goal: Reconstruct y^* given x.

The variational problem

$$\min_y \frac{1}{2}\|y - x\|^2 + R(y) \tag{1}$$

gives stable solutions, where R is known as regularizer.

Bauschke, H. H., Combettes, P. L., 2017.

Inverse problems

Let $\mathcal{X}$ be a separable Hilbert space. The denoising model writes

$$x = y^* + \varepsilon,$$

where x is the data, y^* the ground truth, and ε the noise.

Goal: Reconstruct y^* given x.

The variational problem

$$\min_y \frac{1}{2}\|y - x\|^2 + R(y) \tag{1}$$

gives stable solutions, where R is known as regularizer. If R is proper, convex and lower semicontinuous ($R \in \Gamma_0(\mathcal{X})$), then (1) has a unique minimizer[2]:

$$z = \operatorname*{arg\,min}_{y\in\mathcal{X}} \frac{1}{2}\|y - x\|^2 + R(y)$$

Bauschke, H. H., Combettes, P. L., 2017.

Inverse problems

Let $\mathcal{X}$ be a separable Hilbert space. The denoising model writes

$$x = y^* + \varepsilon,$$

where x is the data, y^* the ground truth, and ε the noise.

Goal: Reconstruct y^* given x.

The variational problem

$$\min_y \frac{1}{2}\|y - x\|^2 + R(y) \tag{1}$$

gives stable solutions, where R is known as regularizer. If R is proper, convex and lower semicontinuous ($R \in \Gamma_0(\mathcal{X})$), then (1) has a unique minimizer[2]:

$$z = \arg\min_{y \in \mathcal{X}} \frac{1}{2}\|y - x\|^2 + R(y) =: \operatorname{prox}_R(x).$$

[2] Bauschke, H. H., Combettes, P. L., 2017.

Some observations

The main problem:

- $\rightsquigarrow$ Given x, which is the best R? Or, “equivalently”, which is the best prox_R?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Some observations

The main problem:

$\rightsquigarrow$ Given x, which is the best R? Or, “equivalently”, which is the best prox_R?

In this talk:

$\rightsquigarrow$ Can we learn prox_R from data?

Some observations

The main problem:

- $\rightsquigarrow$ Given x, which is the best R? Or, “equivalently”, which is the best prox_R?

In this talk:

- $\rightsquigarrow$ Can we learn prox_R from data?

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy data/ground truth term pairs. We want to solve:

$$\min_{T \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \|T(\bar{x}_i) - \bar{y}_i\|^2$$

Some observations

The main problem:

$\rightsquigarrow$ Given x, which is the best R? Or, "equivalently", which is the best prox_R?

In this talk:

$\rightsquigarrow$ Can we learn prox_R from data?

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy data/ground truth term pairs. We want to solve:

$$\min_{T \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \|T(\bar{x}_i) - \bar{y}_i\|^2$$

$\rightsquigarrow$ What is $\mathcal{H}$?

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Some observations

The main part of the problem:

$\rightsquigarrow$ Given x, which is the best R? Or, “equivalently”, which is the best prox_R?

In this talk:

$\rightsquigarrow$ Can we learn prox_R from data?

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy data/ground truth term pairs. We want to solve:

$$\min_{T \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \|T(\bar{x}_i) - \bar{y}_i\|^2$$

$\rightsquigarrow$ What is $\mathcal{H}$?

$\rightsquigarrow$ Can we exploit the structure of prox_R with $R \in \Gamma_0(\mathcal{X})$?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Some observations

The main problem:

- $\rightsquigarrow$ Given x, which is the best R? Or, “equivalently”, which is the best prox_R?

In this talk:

- $\rightsquigarrow$ Can we learn prox_R from data?

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy data/ground truth term pairs. We want to solve:

$$\min_{T \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \|T(\bar{x}_i) - \bar{y}_i\|^2$$

$\rightsquigarrow$ What is $\mathcal{H}$?

$\rightsquigarrow$ Can we exploit the structure of prox_R with $R \in \Gamma_0(\mathcal{X})$?

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Plug-and-Play (PnP) methods[3]

Consider the FB iteration of (1): $x_{k+1} = \mathrm{prox}_R(x_k)$ and generalize it as $x_{k+1} = Tx_k$.

Venkatakrishnan et al., 2013; Ryu, E. et al., 2019; Terris et al., 2021; Hertrich et al. 2021.

Plug-and-Play (PnP) methods[3]

Consider the FB iteration of (1): $x_{k+1} = \mathrm{prox}_R(x_k)$ and generalize it as $x_{k+1} = Tx_k$.

$\rightsquigarrow$ Great experimental results!

$\rightsquigarrow$ Theoretical guarantees?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Venkatakrishnan et al., 2013; Ryu, E. et al., 2019; Terris et al., 2021; Hertrich et al. 2021.

Plug-and-Play (PnP) methods[3]

Consider the FB iteration of (1): $x_{k+1} = \mathrm{prox}_R(x_k)$ and generalize it as $x_{k+1} = Tx_k$.

$\rightsquigarrow$ Great experimental results!
$\rightsquigarrow$ Theoretical guarantees?

How to make $(x_k)_{k\in\mathbb{N}}$ convergent?

[3] Venkatakrishnan et al., 2013; Ryu, E. et al., 2019; Terris et al., 2021; Hertrich et al. 2021.

Plug-and-Play (PnP) methods[3]

Consider the FB iteration of (1): $x_{k+1} = \mathrm{prox}_R(x_k)$ and generalize it as $x_{k+1} = Tx_k$.

$\rightsquigarrow$ Great experimental results!
$\rightsquigarrow$ Theoretical guarantees?

How to make $(x_k)_{k\in\mathbb{N}}$ convergent? FB converges because:

Theorem. Let $T : \mathcal{X} \to \mathcal{X}$ be a firmly nonexpansive operator such that $\mathrm{Fix}\, T \neq \emptyset$. Let $x_0 \in \mathcal{X}$ and set

$$x_{k+1} = Tx_k$$

Then, $x_k - Tx_k \to 0$ and there exists $y \in \mathrm{Fix}\, T$ such that $x_k \rightharpoonup y$.

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

[3] Venkatakrishnan et al., 2013; Ryu, E. et al., 2019; Terris et al., 2021; Hertrich et al. 2021.

Plug-and-Play (PnP) methods[3]

Consider the FB iteration of (1): $x_{k+1} = \mathrm{prox}_R(x_k)$ and generalize it as $x_{k+1} = Tx_k$.

$\rightsquigarrow$ Great experimental results!
$\rightsquigarrow$ Theoretical guarantees?

How to make $(x_k)_{k\in\mathbb{N}}$ convergent? FB converges because:

Theorem. Let $T : \mathcal{X} \to \mathcal{X}$ be a firmly nonexpansive operator such that $\mathrm{Fix}\, T \neq \emptyset$. Let $x_0 \in \mathcal{X}$ and set

$$x_{k+1} = Tx_k$$

Then, $x_k - Tx_k \to 0$ and there exists $y \in \mathrm{Fix}\, T$ such that $x_k \rightharpoonup y$.

$\rightsquigarrow$ prox_R, $R \in \Gamma_0(X)$, is firmly nonexpansive + "firmly nonexpansive $\Leftrightarrow$ nonexpansive."

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

[3] Venkatakrishnan et al., 2013; Ryu, E. et al., 2019; Terris et al., 2021; Hertrich et al. 2021.

Nonexpansive operators

Recall: N nonexpansive if, for every $x, x' \in X$, we have $\|N(x) - N(x')\| \leq \|x - x'\|$.

Nonexpansive operators

Recall: N nonexpansive if, for every $x, x' \in X$, we have $\|N(x) - N(x')\| \leq \|x - x'\|$.

We therefore study

$$\mathcal{N} := \{N : \mathcal{X} \to \mathcal{X} \mid N \text{ is nonexpansive}\}.$$

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Nonexpansive operators

Recall: N nonexpansive if, for every $x, x' \in X$, we have $\|N(x) - N(x')\| \leq \|x - x'\|$.

We therefore study

$$\mathcal{N} := \{N : \mathcal{X} \to \mathcal{X} \mid N \text{ is nonexpansive}\}.$$

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy/clean pairs. We want to solve:

$$\widehat{N} \in \underset{N \in \mathcal{N}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2. \tag{EP}$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Nonexpansive operators

Recall: N nonexpansive if, for every $x, x' \in X$, we have $\|N(x) - N(x')\| \leq \|x - x'\|$.

We therefore study

$$\mathcal{N} := \{N : \mathcal{X} \to \mathcal{X} \mid N \text{ is nonexpansive}\}\,.$$

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy/clean pairs. We want to solve:

$$\widehat{N} \in \underset{N \in \mathcal{N}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2. \tag{EP}$$

$\rightsquigarrow$ (EP) is computationally unfeasible $\rightsquigarrow$ We want to approximate $\mathcal{N}$.

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Nonexpansive operators

Recall: N nonexpansive if, for every $x, x' \in X$, we have $\|N(x) - N(x')\| \leq \|x - x'\|$.

We therefore study

$$\mathcal{N} := \{N : \mathcal{X} \to \mathcal{X} \mid N \text{ is nonexpansive}\}.$$

Consider $\{(\bar{x}_i, \bar{y}_i)\}_{i=1}^n$ set of noisy/clean pairs. We want to solve:

$$\widehat{N} \in \underset{N \in \mathcal{N}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2. \tag{EP}$$

$\rightsquigarrow$ (EP) is computationally unfeasible $\rightsquigarrow$ We want to approximate $\mathcal{N}$.
$\rightsquigarrow$ We propose a "piecewise affine" version of (EP).

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

What do we mean with piecewise affine?

Step 1. Given $D := \{x_i\}_{i=1}^m$, $x_i \in \mathbb{R}^d$, we want to consider a “good” simplicial partition $\mathfrak{T}$ of conv(D);

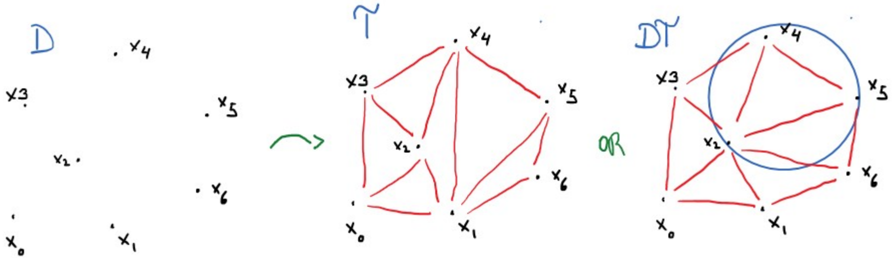
TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

What do we mean with piecewise affine?

Step 1. Given $D := \{x_i\}_{i=1}^m$, $x_i \in \mathbb{R}^d$, we want to consider a "good" simplicial partition $\mathfrak{T}$ of conv(D); i.e, such that

(P1) Every simplex has nonempty interior;

(P2) the intersection of every two simpleces has to be, either empty, or coincide with the convex envelope of its common vertices.



Piecewise affine operators

Step 2. For every $x \in \mathrm{conv}(D)$,

Piecewise affine operators

Step 2. For every $x \in \text{conv}(D)$,there exists a simplex and a collection $\{\lambda_i\}_{i=0}^d$ with $\lambda_i \geq 0$, $\sum_{i=0}^d \lambda_i = 1$, such that

$$x = \sum_{i=0}^{d} \lambda_i x_i$$

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Piecewise affine operators

Step 2. For every $x \in \text{conv}(D)$,there exists a simplex and a collection $\{\lambda_i\}_{i=0}^d$ with $\lambda_i \geq 0$, $\sum_{i=0}^d \lambda_i = 1$, such that

$$x = \sum_{i=0}^{d} \lambda_i x_i \quad \rightsquigarrow \quad N : \text{conv}(D) \to \mathbb{R}^d;\ N(x) := \sum_{i=0}^{d} \lambda_i N(x_i),$$

being $N(x_i) := y_i$, for a given set $\{y_i\}_{i=1}^m$.

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Piecewise affine operators

Step 2. For every $x \in \mathsf{conv}(D)$,there exists a simplex and a collection $\{\lambda_i\}_{i=0}^d$ with $\lambda_i \geq 0$, $\sum_{i=0}^d \lambda_i = 1$, such that

$$x = \sum_{i=0}^{d} \lambda_i x_i \quad \rightsquigarrow \quad N : \mathsf{conv}(D) \to \mathbb{R}^d;\ N(x) := \sum_{i=0}^{d} \lambda_i N(x_i),$$

being $N(x_i) := y_i$, for a given set $\{y_i\}_{i=1}^m$.

Define

$$\mathsf{PA}(\mathfrak{T}) := \left\{ N : \mathsf{conv}(D) \to \mathbb{R}^d \ :\ N(x) := \sum_{i=0}^{d} \lambda_i N(x_i) \right\}.$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019 TraDE-OPT

The piecewise affine problem

Our piecewise affine problem:

$$\min_{N \in \mathcal{N} \cap \mathsf{PA}(\mathfrak{T})} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2. \tag{PAP}$$

The piecewise affine problem

Our piecewise affine problem:

$$\min_{N \in \mathcal{N} \cap \mathsf{PA}(\mathfrak{T})} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2. \qquad \text{(PAP)}$$

Some observations:

- (PAP) is not a computational-friendly formulation.

The piecewise affine problem

Our piecewise affine problem:

$$\min_{N \in \mathcal{N} \cap \mathsf{PA}(\mathfrak{T})} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2. \tag{PAP}$$

Some observations:

- (PAP) is not a computational-friendly formulation.
- Can we find a condition in $\{(\bar{x}_i, N(\bar{x}_i))\}_{i=1}^{n}$ so that the extended operator (in the piecewise affine sense) is 1-Lipschitz?

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Seeking for a condition

Every $x \in \text{conv}(D)$ can be rewritten as

$$x = \sum_{i=0}^{d} \lambda_i x_i = \sum_{i=0}^{d} \lambda_i (x_i - x_0) + x_0,$$

for some simplex,

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Seeking for a condition

Every $x \in \text{conv}(D)$ can be rewritten as

$$x = \sum_{i=0}^{d} \lambda_i x_i = \sum_{i=0}^{d} \lambda_i (x_i - x_0) + x_0,$$

for some simplex, with image

$$N(x) := \sum_{i=1}^{n} \lambda_i y_i = BA^{-1}(x - x_0) + y_0,$$

where $A = (x_1 - x_0 \mid ... \mid x_d - x_0)$, and $B = (y_1 - y_0 \mid ... \mid y_d - y_0)$.

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

We found it!

Theorem

If, for every simplex $\|BA^{-1}\| \leq 1$, then N (as above) is nonexpansive.

We found it!

Theorem

If, for every simplex $\|BA^{-1}\| \leq 1$, then N (as above) is nonexpansive.

Consider

$$\min_{y_1,\dots y_m} \frac{1}{n} \sum_{i=1}^{n} \|y_i - \bar{y}_i\|^2 \quad \text{s.t. } \|B(y_1, \dots y_m)A^{-1}\| \leq 1, \text{ for every simplex.} \tag{FP}$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

We found it!

Theorem

If, for every simplex $\|BA^{-1}\| \leq 1$, then N (as above) is nonexpansive.

Consider

$$\begin{aligned} &\min_{y_1,\dots y_m} \frac{1}{n}\sum_{i=1}^{n} \|y_i - \bar{y}_i\|^2 \\ &\text{s.t. } \|B(y_1,\dots y_m)A^{-1}\| \leq 1, \text{ for every simplex.} \end{aligned} \tag{FP}$$

Then,

Theorem

Problems (FP) and (PAP) are equivalent.

A convergence result

Can we prove that, in some sense, that (PAP) converges to (EP) at least in conv(D)?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

A convergence result

Can we prove that, in some sense, that (PAP) converges to (EP) at least in conv(D)?

Theorem

For every $k \in \mathbb{N}$, there exists a triangulation $\mathfrak{T}_k$

A convergence result

Can we prove that, in some sense, that (PAP) converges to (EP) at least in conv(D)?

Theorem

For every $k \in \mathbb{N}$, there exists a triangulation $\mathfrak{T}_k$ such that, if we define

$$\widehat{N}_k \in \underset{L \in \mathsf{PA}(\mathfrak{T}_k) \cap \mathcal{N}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|N_k(\bar{x}_i) - \bar{y}_i\|^2,$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

A convergence result

Can we prove that, in some sense, that (PAP) converges to (EP) at least in conv(D)?

Theorem

For every $k \in \mathbb{N}$, there exists a triangulation $\mathfrak{T}_k$ such that, if we define

$$\widehat{N}_k \in \underset{L \in \mathsf{PA}(\mathfrak{T}_k) \cap \mathcal{N}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|N_k(\bar{x}_i) - \bar{y}_i\|^2,$$

there exists a subsequence $\{k_j\}_{j=1}^{\infty}$ such that $\widehat{N}_{k_j} \overset{*}{\rightharpoonup} \widehat{N}$, being

$$\widehat{N} \in \underset{N \in \mathcal{N}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{y}_i\|^2.$$

Experiments

In practice, we search for

$$\arg\min_{x\in\mathbb{R}^{n\times n}} \frac{1}{2\alpha}\|x-y\|^2 + G(Dx),$$

where $(Dx)_i = (x_{i+1,j} - x_{i,j},\, x_{i,j+1} - x_{i,j})$, $i = 1, ..., n^2$. To solve the dimensionality problem of simplicial partitions, we suppose G is of the form

$$G(v) = \sum_{i=1}^{n^2} G_i(v_i), \quad \text{for every } v \in \mathbb{R}^{n^2\times 2},$$

where $G_i : \mathbb{R}^2 \to (-\infty, +\infty]$. With this, we learn $\text{prox}_{G_i} : \mathbb{R}^2 \to \mathbb{R}^2$.

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Experiments

In practice, we search for

$$\underset{x\in\mathbb{R}^{n\times n}}{\arg\min}\ \frac{1}{2\alpha}\|x-y\|^2+G(Dx),$$

where $(Dx)_i=(x_{i+1,j}-x_{i,j},\, x_{i,j+1}-x_{i,j})$, $i=1,...,n^2$. To solve the dimensionality problem of simplicial partitions, we suppose G is of the form

$$G(v)=\sum_{i=1}^{n^2}G_i(v_i),\qquad \text{for every } v\in\mathbb{R}^{n^2\times 2},$$

where $G_i:\mathbb{R}^2\to(-\infty,+\infty]$. With this, we learn $\text{prox}_{G_i}:\mathbb{R}^2\to\mathbb{R}^2$.

Example: If $G=\|\cdot\|^2$, $\|Dx\|^2=\sum_{i=1}^{n^2}\|(Dx)_i\|^2$ and so

$$\text{prox}_{\|\cdot\|^2}(x_1,...,x_{n^2})=(\text{prox}_{\|\cdot\|^2}(x_1),...,\text{prox}_{\|\cdot\|^2}(x_{n^2})).$$

Some pictures

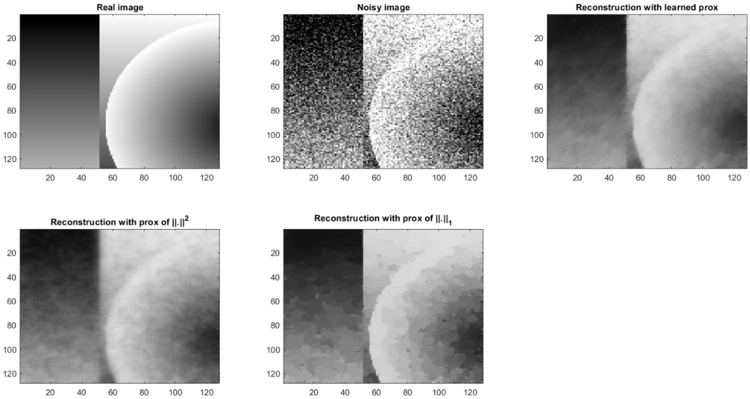


Figure: Classic regularizers compared to ours.

Conclusions

- We provide a constructive method to learn nonexpansive (and therefore also firmly nonexpansive) operators,
- classic optim. algorithms such as Chambolle–Pock, Douglas–Rachford, Forward-Backward Splitting or ADMM have their corresponding convergent PnP version,
- we prove that the approximate problem that we define converges to the original empirical one,
- preprint will be soon (I hope) on arXiv!

TraDE-OPT TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

¡Muchas gracias!

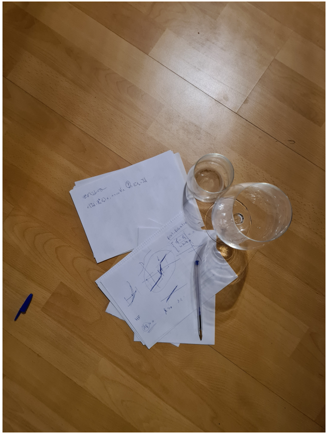


Figure: vino y teoremas

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019