TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Learning firmly nonexpansive operators

SIGMA Conference, CIRM

Jonathan Chirinos Rodríguez

Joint work with K. Bredies and E. Naldi

(https://arxiv.org/abs/2407.14156)

31^{st} of October, 2024

Institut de Recherche en Informatique de Toulouse
CNRS - Toulouse INP - UT3 - UT Capitole - UT2

UniGe | DIMA

Outline

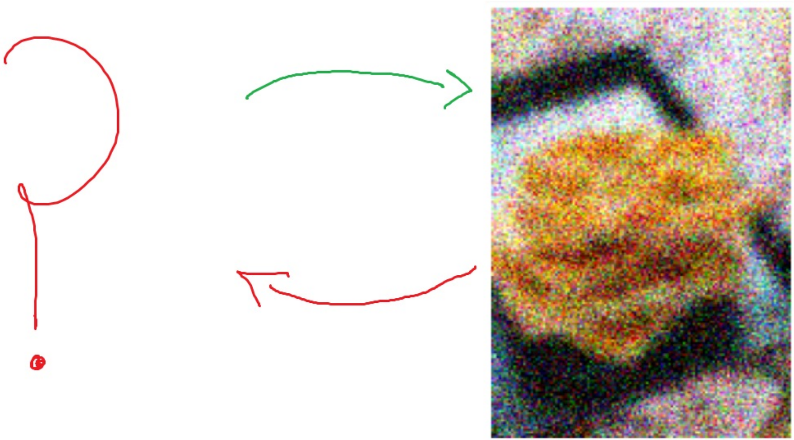
Background

Setting and results

Applications

TraDE-OPT
TRAINING DATA-DRIVEN EXPERTS IN
O P T I M I Z A T I O N
M S C A - I T N 2 0 1 9

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019



Inverse problems

Let $u^* \in \mathcal{U}$, $x \in \mathcal{X}$ and $A : \mathcal{U} \to \mathcal{X}$ an operator. Consider

$$x = Au^* + \varepsilon, \quad \varepsilon \text{ noise.}$$

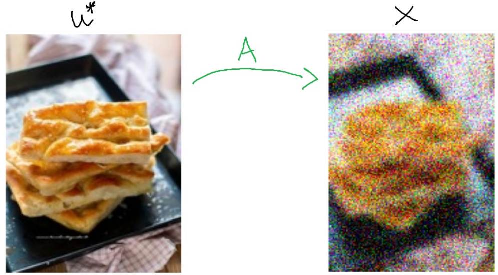
TraDE-OPT
TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION
MSCA-ITN 2019

Inverse problems

Let $u^* \in \mathcal{U}$, $x \in \mathcal{X}$ and $A : \mathcal{U} \to \mathcal{X}$ an operator. Consider

$$x = Au^* + \varepsilon, \quad \varepsilon \text{ noise.}$$

Goal: Reconstruct u^* given x.

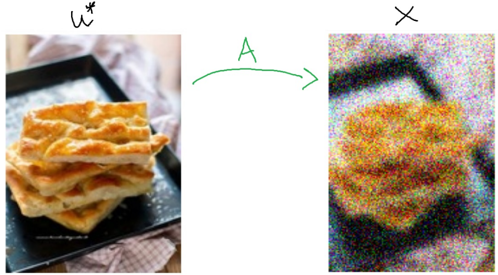


Inverse problems

Let $u^* \in \mathcal{U}$, $x \in \mathcal{X}$ and $A : \mathcal{U} \to \mathcal{X}$ an operator. Consider

$$x = Au^* + \varepsilon, \quad \varepsilon \text{ noise.}$$

Goal: Reconstruct u^* given x.



Examples of A:

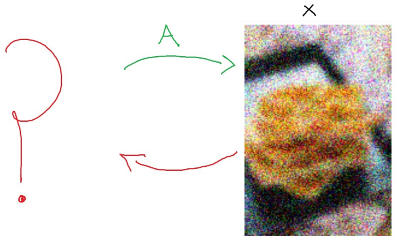
- $\rightsquigarrow$ Denoising: $A = \mathrm{Id}$ identity map
- $\rightsquigarrow$ Deblurring: $Au = \kappa * u$ convolution operator
- $\rightsquigarrow$ Phase Retrieval: $A(u) = |\mathcal{F}u|$ modulus of the Fourier transform

Ill-posedness

Let $u^* \in \mathcal{U}$, $x \in \mathcal{X}$ and $A : \mathcal{U} \to \mathcal{X}$ an operator. Consider

$$x = Au^* + \varepsilon, \quad \varepsilon \text{ noise}.$$

Goal: Reconstruct u^* given x.



Problem: Ill-posedness![1]

[1] Engl, H. W. et al., 1996.

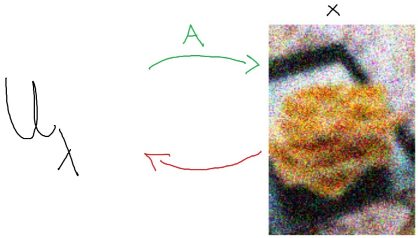
TraDE-OPT
TRAINING DATA-DRIVEN EXPERTS IN
OPTIMIZATION
MSCA-ITN 2019

Two main approaches

TRAINING DATA-DRIVEN EXPERTS IN
OPTIMIZATION
MSCA-ITN 2019

Two main approaches

Model-based

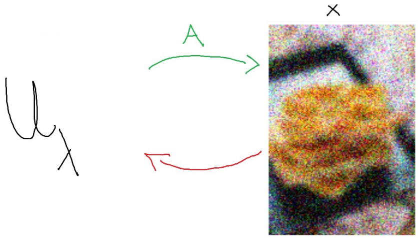


$$u_\lambda \in \arg\min_u \ell(Au, x) + \lambda R(u)$$

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

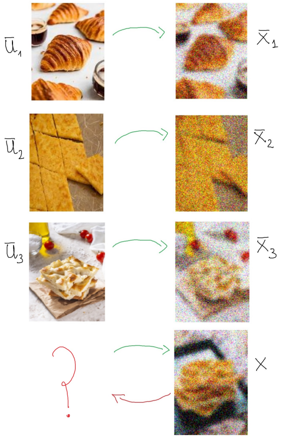
Two main approaches

Model-based

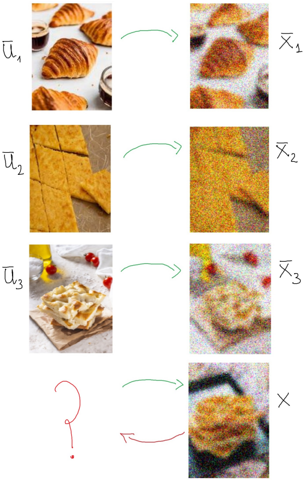


$$u_\lambda \in \arg\min_u \ell(Au, x) + \lambda R(u)$$

Data-driven

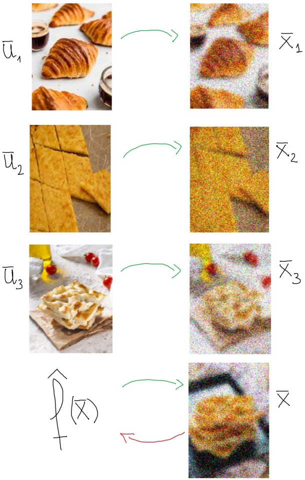


2.- Data-driven approaches



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

2.- Data-driven approaches[2]

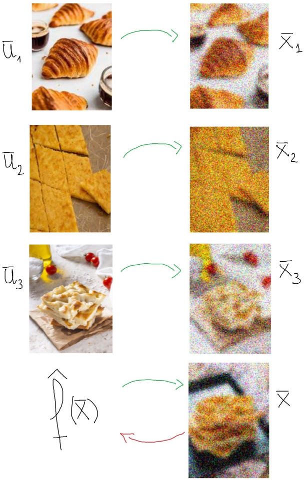


Find a map

$$((\bar{X}_1, \bar{U}_1), \dots, (\bar{X}_n, \bar{U}_n)) \mapsto \widehat{f}.$$

[2]Arridge et al, 2019.

2.- Data-driven approaches[2]



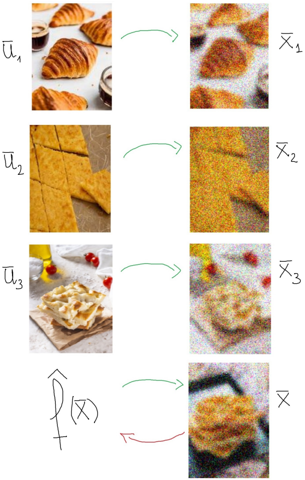
Find a map

$$((\bar{X}_1, \bar{U}_1), \ldots, (\bar{X}_n, \bar{U}_n)) \mapsto \widehat{f}.$$

$\rightsquigarrow$ Given a **new** $\bar{X}$, $\widehat{f}(\bar{X}) \approx \bar{U}^*$.

[2] Arridge et al, 2019.

2.- Data-driven approaches[2]



Find a map

$$((\bar{X}_1, \bar{U}_1), \ldots, (\bar{X}_n, \bar{U}_n)) \mapsto \hat{f}.$$

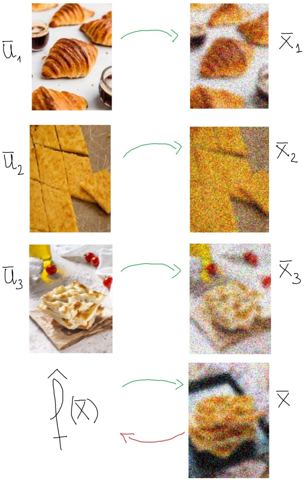
$\rightsquigarrow$ Given a **new** $\bar{X}$, $\hat{f}(\bar{X}) \approx \bar{U}^*$.

Two types of $\hat{f}$:

1. $\hat{f}$ is agnostic/black-box,

[2] Arridge et al, 2019.

2.- Data-driven approaches[2]



Find a map

$$((\bar{X}_1, \bar{U}_1), \ldots, (\bar{X}_n, \bar{U}_n)) \mapsto \widehat{f}.$$

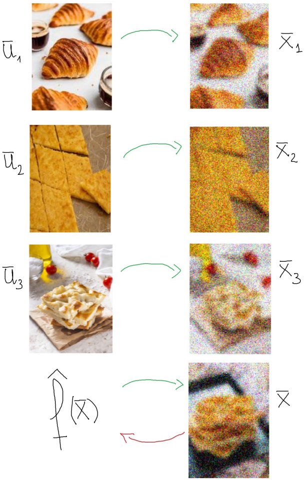
$\rightsquigarrow$ Given a **new** $\bar{X}$, $\widehat{f}(\bar{X}) \approx \bar{U}^*$.

Two types of $\widehat{f}$:

1. $\widehat{f}$ is agnostic/black-box,
2. $\widehat{f} = \widehat{f}(A, \ldots)$.

[2]Arridge et al, 2019.

2.- Data-driven approaches[2]



Find a map

$$((\bar{X}_1, \bar{U}_1), \ldots, (\bar{X}_n, \bar{U}_n)) \mapsto \hat{f}.$$

$\rightsquigarrow$ Given a **new** $\bar{X}$, $\hat{f}(\bar{X}) \approx \bar{U}^*$.

Two types of $\hat{f}$:

1. $\hat{f}$ is agnostic/black-box,
2. $\hat{f} = \hat{f}(A, \ldots)$.

Great results in practice!!

Theory ?

[2]Arridge et al, 2019.

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Today: learning R

Today: learning R

$\rightsquigarrow$ Take $A = \mathrm{Id}$ (denoising).

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Today: learning R

$\rightsquigarrow$ Take $A = \mathrm{Id}$ (denoising). Learning $R \in \Gamma_0(\mathcal{U})$ is **equivalent** to learning

$$\bar{U}_R = \arg\min_u \frac{1}{2}\|u - \bar{X}\|^2 + R(u) = \mathrm{prox}_R(\bar{X}).$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Today: learning R

$\rightsquigarrow$ Take $A = \mathrm{Id}$ (denoising). Learning $R \in \Gamma_0(\mathcal{U})$ is **equivalent** to learning

$$\bar{U}_R = \arg\min_u \frac{1}{2}\|u - \bar{X}\|^2 + R(u) = \mathrm{prox}_R(\bar{X}).$$

where, for every $x \in \mathcal{X}$,

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

$\rightsquigarrow$ Given R, we fix, for every $i = 1, \ldots, n$,

$$\bar{U}_R^i = \mathrm{prox}_R(\bar{X}_i).$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Today: learning R

$\rightsquigarrow$ Take $A = \mathrm{Id}$ (denoising). Learning $R \in \Gamma_0(\mathcal{U})$ is **equivalent** to learning

$$\bar{U}_R = \arg\min_u \frac{1}{2}\|u - \bar{X}\|^2 + R(u) = \mathrm{prox}_R(\bar{X}).$$

where, for every $x \in \mathcal{X}$,

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

$\rightsquigarrow$ Given R, we fix, for every $i = 1, \ldots, n$,

$$\bar{U}_R^i = \mathrm{prox}_R(\bar{X}_i).$$

$\rightsquigarrow$ Then, we learn R:

$$\widehat{R} \in \arg\min_{R \in \mathcal{R}} \frac{1}{2}\sum_{i=1}^{n} \|\bar{U}_R^i - \bar{U}_i\|^2.$$

TraDE-OPT
TRAINING DATA-DRIVEN EXPERTS IN
O P T I M I Z A T I O N
M S C A - I T N 2 0 1 9

Outline

Background

Setting and results

Applications

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

Setting: denoising and supervised learning

$\rightsquigarrow$ Let $\mathcal{X} = \mathcal{U}$ and $A = \mathrm{Id}$. Consider

$$\bar{X} = \bar{U} + \varepsilon,$$

where $\bar{X}$, $\bar{U}$ and ε are random variables.

Setting: denoising and supervised learning

$\rightsquigarrow$ Let $\mathcal{X} = \mathcal{U}$ and $A = \mathrm{Id}$. Consider

$$\bar{X} = \bar{U} + \varepsilon,$$

where $\bar{X}$, $\bar{U}$ and ε are random variables.

$\rightsquigarrow$ We **don't** have access to the prob. distribution of $(\bar{X}, \bar{U})$,

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Setting: denoising and supervised learning

$\rightsquigarrow$ Let $\mathcal{X} = \mathcal{U}$ and $A = \mathrm{Id}$. Consider

$$\bar{X} = \bar{U} + \varepsilon,$$

where $\bar{X}$, $\bar{U}$ and ε are random variables.

$\rightsquigarrow$ We **don't** have access to the prob. distribution of $(\bar{X}, \bar{U})$,

$\rightsquigarrow$ **BUT**

1. We assume

$$\mathbb{E}[\|\bar{X}\|^2 + \|\bar{U}\|^2] < +\infty.$$

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

Setting: denoising and supervised learning

$\rightsquigarrow$ Let $\mathcal{X} = \mathcal{U}$ and $A = \mathrm{Id}$. Consider

$$\bar{X} = \bar{U} + \varepsilon,$$

where $\bar{X}$, $\bar{U}$ and ε are random variables.

$\rightsquigarrow$ We **don't** have access to the prob. distribution of $(\bar{X}, \bar{U})$,

$\rightsquigarrow$ **BUT**

1. We assume

$$\mathbb{E}[\|\bar{X}\|^2 + \|\bar{U}\|^2] < +\infty.$$

2. We have access to n independent and identical copies $\{(\bar{X}_i, \bar{U}_i)\}_{i=1}^n$ of $(\bar{X}, \bar{U})$.

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

Bilevel approach

For every $i = 1, \dots, n$, fix

$$\bar{U}_R^i = \operatorname{prox}_R(\bar{X}_i).$$

Then, we learn R:

$$\widehat{R} \in \operatorname*{arg\,min}_{R \in \mathcal{R}} \frac{1}{2} \sum_{i=1}^{n} \|\bar{U}_R^i - \bar{U}_i\|^2.$$

TraDE-OPT

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Bilevel approach

For every $i = 1, \ldots, n$, fix

$$\bar{U}_R^i = \operatorname{prox}_R(\bar{X}_i).$$

Then, we learn R:

$$\widehat{R} \in \underset{R \in \mathcal{R}}{\arg\min} \; \frac{1}{2} \sum_{i=1}^{n} \|\bar{U}_R^i - \bar{U}_i\|^2.$$

BUT:

1. What is a reasonable choice for $\mathcal{R}$?

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

Bilevel approach

For every $i = 1, \ldots, n$, fix

$$\bar{U}_R^i = \mathrm{prox}_R(\bar{X}_i).$$

Then, we learn R:

$$\widehat{R} \in \operatorname*{arg\,min}_{R \in \mathcal{R}} \frac{1}{2} \sum_{i=1}^{n} \|\bar{U}_R^i - \bar{U}_i\|^2.$$

BUT:

1. What is a reasonable choice for $\mathcal{R}$?
2. How to solve the bilevel problem?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Computational intractability

$\rightsquigarrow \mathcal{R} = \Gamma_0(\mathcal{X})$ unstructured set!

TRAINING DATA-DRIVEN EXPERTS IN
OPTIMIZATION
MSCA-ITN 2019

Computational intractability

$\rightsquigarrow \mathcal{R} = \Gamma_0(\mathcal{X})$ unstructured set!

Possible approaches:

1. Parametrized[3] R (e.g. $\widehat{R}(u) = \|\widehat{B}^{-1}(u - \widehat{h})\|^2$),

[3] Alberti, De Vito, Lassas, Ratti, Santacesaria, 2021; Pock et al, 2020.

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

Computational intractability

$\rightsquigarrow \mathcal{R} = \Gamma_0(\mathcal{X})$ unstructured set!

Possible approaches:

1. Parametrized[3] R (e.g. $\widehat{R}(u) = \|\widehat{B}^{-1}(u - \widehat{h})\|^2$),
2. **"Relaxation"** of prox_R:

$$\{\mathrm{prox}_R \mid R \in \Gamma_0(\mathcal{X})\} \subset \mathcal{N} := \{N : \mathcal{X} \to \mathcal{X} \mid N \text{ is nonexpansive}\}$$

[3] Alberti, De Vito, Lassas, Ratti, Santacesaria, 2021; Pock et al, 2020.

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

Expected/Empirical Risk Minimization

For every $i = 1, \dots, n$, fix

$$\bar{U}_R(\bar{X}_i) := N(\bar{X}_i).$$

Then, we learn N:

$$N_n^* \in \operatorname*{arg\,min}_{N \in \mathcal{N}} \; L_n(N) := \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{X}_i) - \bar{U}_i\|^2.$$

TRAINING DATA-DRIVEN EXPERTS IN
OPTIMIZATION
MSCA-ITN 2019

Expected/Empirical Risk Minimization

For every $i = 1, \dots, n$, fix

$$\bar{U}_R(\bar{X}_i) := N(\bar{X}_i).$$

Then, we learn N:

$$N_n^* \in \underset{N \in \mathcal{N}}{\arg\min}\ L_n(N) := \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{X}_i) - \bar{U}_i\|^2. \qquad \text{(EP)}$$

QUESTION: Given

$$N^* \in \underset{N \in \mathcal{N}}{\arg\min}\ L(N) := \mathbb{E}[\|N(\bar{X}) - \bar{U}\|^2],$$

Expected/Empirical Risk Minimization

For every $i = 1, \ldots, n$, fix

$$\bar{U}_R(\bar{X}_i) := N(\bar{X}_i).$$

Then, we learn N:

$$N_n^* \in \operatorname*{arg\,min}_{N \in \mathcal{N}} L_n(N) := \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{X}_i) - \bar{U}_i\|^2. \tag{EP}$$

QUESTION: Given

$$N^* \in \operatorname*{arg\,min}_{N \in \mathcal{N}} L(N) := \mathbb{E}[\|N(\bar{X}) - \bar{U}\|^2], \tag{CP}$$

can we show that N_n^* is a good approximation of N^*?

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

Expected/Empirical Risk Minimization

For every $i = 1, \dots, n$, fix

$$\bar{U}_R(\bar{X}_i) := N(\bar{X}_i).$$

Then, we learn N:

$$N_n^* \in \arg\min_{N \in \mathcal{N}} L_n(N) := \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{X}_i) - \bar{U}_i\|^2. \qquad \text{(EP)}$$

QUESTION: Given

$$N^* \in \arg\min_{N \in \mathcal{N}} L(N) := \mathbb{E}[\|N(\bar{X}) - \bar{U}\|^2], \qquad \text{(CP)}$$

can we show that N_n^* is a good approximation of N^*?

(**NOTE:** $\mathbb{E}[\|\bar{X}\|^2 + \|\bar{U}\|^2] < +\infty \implies N^*$ exists!)

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

A Gamma convergence result

Theorem (Bredies, CR, Naldi)

L_n Γ-converges to L almost surely as $n \to \infty$.

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

A Gamma convergence result

Theorem (Bredies, CR, Naldi)

$$L_n \ \Gamma\text{-converges to } L \text{ almost surely as } n \to \infty.$$

Corollary

Let $(N_n^*)_{n\in\mathbb{N}}$, be the sequence of minimizers of L_n for every $n \in \mathbb{N}$. Then, there exists a minimizer N^* of L such that, up to subsequences,

$$N_n^* \overset{*}{\rightharpoonup} N^*, \quad \text{a.s., as } n \to \infty.$$

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

Can we do better?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION
MSCA-ITN 2019

Can we do better?

RECALL: prox_R, $R \in \Gamma_0(\mathcal{X})$, is a firmly nonexpansive (FNE) operator.

TRAINING DATA-DRIVEN EXPERTS IN
OPTIMIZATION
MSCA-ITN 2019

Can we do better?

RECALL: prox_R, $R \in \Gamma_0(\mathcal{X})$, is a firmly nonexpansive (FNE) operator.

FACT: For every $T \colon \mathcal{X} \to \mathcal{X}$ FNE, there exists $N \colon \mathcal{X} \to \mathcal{X}$ NE such that

$$T = \frac{1}{2}\mathrm{Id} + \frac{1}{2}N,$$

and viceversa.

Can we do better?

RECALL: prox_R, $R \in \Gamma_0(\mathcal{X})$, is a firmly nonexpansive (FNE) operator.

FACT: For every $T\colon \mathcal{X} \to \mathcal{X}$ FNE, there exists $N\colon \mathcal{X} \to \mathcal{X}$ NE such that

$$T = \frac{1}{2}\mathrm{Id} + \frac{1}{2}N,$$

and viceversa. Therefore, with $\bar{U}_i' := \frac{1}{2}(\bar{U}_i + \bar{X}_i)$,

$$\operatorname*{arg\,min}_{N \in \mathcal{N}} \frac{1}{n}\sum_{i=1}^{n} \|N(\bar{X}_i) - \bar{U}_i\|^2 \quad \text{and} \quad \operatorname*{arg\,min}_{T \text{ is FNE}} \frac{1}{n}\sum_{i=1}^{n} \|T(\bar{X}_i) - \bar{U}_i'\|^2$$

are equivalent!

(Note: $\arg\min_T \ \|T(x) - u'\|^2 \ \equiv \ \arg\min_T \ 4\|T(x) - u'\|^2 \ \equiv \ \arg\min_N \ \|N(x) - u\|^2$).

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

How to solve (EP) in practice?

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

How to solve (EP) in practice?

$\rightsquigarrow$ **Fix** $\mathcal{X} = \mathbb{R}^d$, $d \geq 2$,

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

How to solve (EP) in practice?

$\rightsquigarrow$ **Fix** $\mathcal{X} = \mathbb{R}^d$, $d \geq 2$, consider finitely many samples $(\bar{x}_1, \bar{u}_1), \ldots, (\bar{x}_n, \bar{u}_n)$,

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

How to solve (EP) in practice?

$\rightsquigarrow$ **Fix** $\mathcal{X} = \mathbb{R}^d$, $d \geq 2$, consider finitely many samples $(\bar{x}_1, \bar{u}_1), \dots, (\bar{x}_n, \bar{u}_n)$, and solve

$$\widehat{N} \in \arg\min_{N \in \mathcal{N}} \widehat{L}(N) := \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{u}_i\|^2.$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

How to solve (EP) in practice?

$\rightsquigarrow$ **Fix** $\mathcal{X} = \mathbb{R}^d$, $d \geq 2$, consider finitely many samples $(\bar{x}_1, \bar{u}_1), \ldots, (\bar{x}_n, \bar{u}_n)$, and solve

$$\widehat{N} \in \arg\min_{N \in \mathcal{N}} \widehat{L}(N) := \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{u}_i\|^2. \qquad \text{(DP)}$$

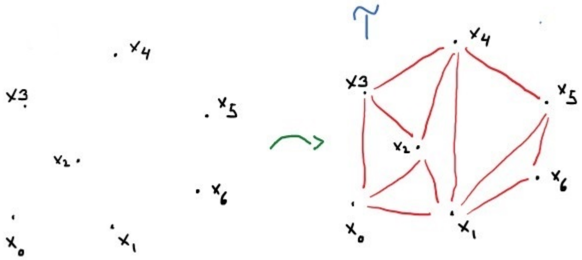
$\rightsquigarrow$ **BUT:** $\mathcal{N}$ is infinite-dimensional $\rightsquigarrow$ We need to discretize it.

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

Simplicial partitions

Construct a “piecewise affine” approximation of $\mathcal{N}$.

Step 1. Given $D := \{x_1, \dots, x_m\}$, $x_i \in \mathbb{R}^d$, $m \geq d+1$, let $\mathfrak{T}$ be a simplicial partition of conv(D) such that



Piecewise affine operators: part I

Step 2. Denote $\mathfrak{T} := \{S_1, \dots, S_\ell\}$, $\ell \in \mathbb{N}$, and consider

$$\lambda_1, \dots, \lambda_m \colon \text{conv}(D) \to [0,1]$$

the **Lagrange elements** of order 1 associated with $\mathfrak{T}$; i.e. such that

(i) $\lambda_i(x_j) = \delta_{ij}$, for $i,\ j = 1, \dots, m$, $\quad \delta_{ij}$ Kronecker delta,

(ii) $\lambda_i|_{S_t}$ is a polynomial of degree ≤ 1 for each $i = 1, \dots, m$ and $t = 1, \dots, \ell$.

For every $t = 1, \dots, \ell$, denote $i_0, \dots, i_d$ the indices of the vertices that form S_t. We have

$$\sum_{j=0}^{d} \lambda_{i_j}(x) = 1, \quad \sum_{j=0}^{d} \lambda_{i_j}(x) x_{i_j} = x.$$

Finally, for any $x \in \mathrm{conv}(D)$,

$$\sum_{i=1}^{m} \lambda_i = \chi_{\text{conv}(D)}, \quad \text{and} \quad \sum_{i=1}^{m} \lambda_i x_i = \mathrm{Id}|_{\text{conv}(D)}.$$

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

Piecewise affine operators: part II

Step 3. Given $D' = \{u_1, \ldots, u_m\}$, define

$$\widetilde{N}\colon \mathrm{conv}(D) \to \mathbb{R}^d; \quad \widetilde{N}(x) := \sum_{i=1}^{m} \lambda_i(x) u_i.$$

Then, $N := \tilde{N} \circ \pi_{\mathrm{conv}(D)} : \mathbb{R}^d \to \mathbb{R}^d$. **Note:** N is nonexpansive!

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

The piecewise affine problem

Step 4: Define

$$\mathrm{PA}(\mathfrak{T}) := \left\{ N : \mathbb{R}^d \to \mathbb{R}^d \,|\, N := \tilde{N} \circ \pi_{\mathsf{conv}(D)} \right\}.$$

Finally,

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

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

The piecewise affine problem

Step 4: Define

$$\mathrm{PA}(\mathfrak{T}) := \left\{ N : \mathbb{R}^d \to \mathbb{R}^d \,|\, N := \tilde{N} \circ \pi_{\mathsf{conv}(D)} \right\}.$$

Finally,

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

(We provide a computational-friendly formulation for (PAP) !!)

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

A convergence result: (PAP) to (DP)

Theorem (Bredies, CR, Naldi)
Let $(\mathfrak{T}_k)_k$ be a sequence of "regular" simplicial partitions for conv(D).

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

A convergence result: (PAP) to (DP)

Theorem (Bredies, CR, Naldi)
Let $(\mathfrak{T}_k)_k$ be a sequence of "regular" simplicial partitions for $\operatorname{conv}(D)$. If we define

$$\widehat{N}_k \in \operatorname*{arg\,min}_{N \in \mathrm{PA}(\mathfrak{T}_k) \cap \mathcal{N}} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{u}_i\|^2,$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION
MSCA-ITN 2019

A convergence result: (PAP) to (DP)

Theorem **(Bredies, CR, Naldi)**
Let $(\mathfrak{T}_k)_k$ be a sequence of "regular" simplicial partitions for conv(D). If we define

$$\widehat{N}_k \in \operatorname*{arg\,min}_{N \in \mathrm{PA}(\mathfrak{T}_k) \cap \mathcal{N}} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{u}_i\|^2,$$

then, up to subsequences, $\widehat{N}_k \overset{*}{\rightharpoonup} \widehat{N}$, being

$$\widehat{N} \in \operatorname*{arg\,min}_{N \in \mathcal{N}} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{u}_i\|^2.$$

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

Outline

Background

Setting and results

Applications

An application

Let $A\colon \mathcal{U} \to \mathcal{X}$ be a linear operator. Consider

$$x = Au^* + \varepsilon.$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

An application

Let $A\colon \mathcal{U} \to \mathcal{X}$ be a linear operator. Consider

$$x = Au^* + \varepsilon.$$

The Forward-Backward iteration of

$$\min_u \frac{1}{2}\|Au - x\|^2 + R(u) \tag{var}$$

reads as

$$u_{k+1} = \mathrm{prox}_R(u_k - \tau A^*(Au_k - x)),$$

for some stepsize $\tau > 0$.

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

PnP methods[4]

PnP methods: Substitute prox_R with P acting as a denoiser:

$$u_{k+1} = P(u_k - \tau A^*(Au_k - x)).$$

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

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

PnP methods[4]

PnP methods: Substitute prox_R with P acting as a denoiser:

$$u_{k+1} = P(u_k - \tau A^*(Au_k - x)). \tag{1}$$

$\rightsquigarrow$ + algorithms (CP, DR, ADMM,...).

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

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

PnP methods[4]

PnP methods: Substitute prox_R with P acting as a denoiser:

$$u_{k+1} = P(u_k - \tau A^*(Au_k - x)). \tag{1}$$

$\rightsquigarrow$ + algorithms (CP, DR, ADMM,...).

Few theoretical guarantees: does (1) converge to a solution of (var)/fixed point of P?

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

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

PnP methods[4]

PnP methods: Substitute prox_R with P acting as a denoiser:

$$u_{k+1} = P(u_k - \tau A^*(Au_k - x)). \tag{1}$$

$\rightsquigarrow$ + algorithms (CP, DR, ADMM,...).

Few theoretical guarantees: does (1) converge to a solution of (var)/fixed point of P?

In our case: Substitute prox_R with $\widehat{T} := \frac{1}{2}\text{Id} + \frac{1}{2}\widehat{N}$:

$$u_{k+1} = \widehat{T}(u_k - \tau A^*(Au_k - x)).$$

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

PnP for image denoising

Let $\mathcal{X} = \mathbb{R}^{N \times N}$. Given

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

with $\varepsilon \sim N(0, \tau^2 \mathrm{Id})$, we search for

$$\arg\min_u \|u - x\|_F^2 + R(Du).$$

TRAINING DATA-DRIVEN EXPERTS IN OPTIMIZATION MSCA-ITN 2019

PnP for image denoising

Let $\mathcal{X} = \mathbb{R}^{N\times N}$. Given

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

with $\varepsilon \sim N(0, \tau^2 \mathrm{Id})$, we search for

$$\arg\min_u \|u - x\|_F^2 + R(Du).$$

We want to learn prox_R.

TRAINING DATA-DRIVEN EXPERTS IN
OPTIMIZATION
MSCA-ITN 2019

PnP for image denoising

Let $\mathcal{X} = \mathbb{R}^{N\times N}$. Given

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

with $\varepsilon \sim N(0, \tau^2 \mathrm{Id})$, we search for

$$\arg\min_u \|u - x\|_F^2 + R(Du).$$

We want to learn prox_R. We first suppose

$$R(v) = \sum_{i,j=1}^{N} r(v_{i,j}),$$

where $r : \mathbb{R}^2 \to (-\infty, +\infty]$. With this, we learn $\mathrm{prox}_r : \mathbb{R}^2 \to \mathbb{R}^2$.

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

PnP for image denoising

Let $\mathcal{X} = \mathbb{R}^{N \times N}$. Given

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

with $\varepsilon \sim N(0, \tau^2 \mathrm{Id})$, we search for

$$\arg\min_u \|u - x\|_F^2 + R(Du).$$

We want to learn prox_R. We first suppose

$$R(v) = \sum_{i,j=1}^{N} r(v_{i,j}),$$

where $r : \mathbb{R}^2 \to (-\infty, +\infty]$. With this, we learn $\mathrm{prox}_r : \mathbb{R}^2 \to \mathbb{R}^2$.

Example: If $R = \|\cdot\|_{1,1}$, recall $\|Du\|_{1,1} = \sum_{i,j=1}^{N} \|(D_v u, D_h u)_{i,j}\|_1$, and so, if $v = (v_1, v_2) \in \mathbb{R}^{2 \times N^2}$,

$$\mathrm{prox}_{\|\cdot\|_{1,1}}(v) = (\mathrm{prox}_{\|\cdot\|_1}(v_{1,i}), \mathrm{prox}_{\|\cdot\|_1}(v_{2,i}))_{i=1}^{N^2}.$$

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

Results on the circle

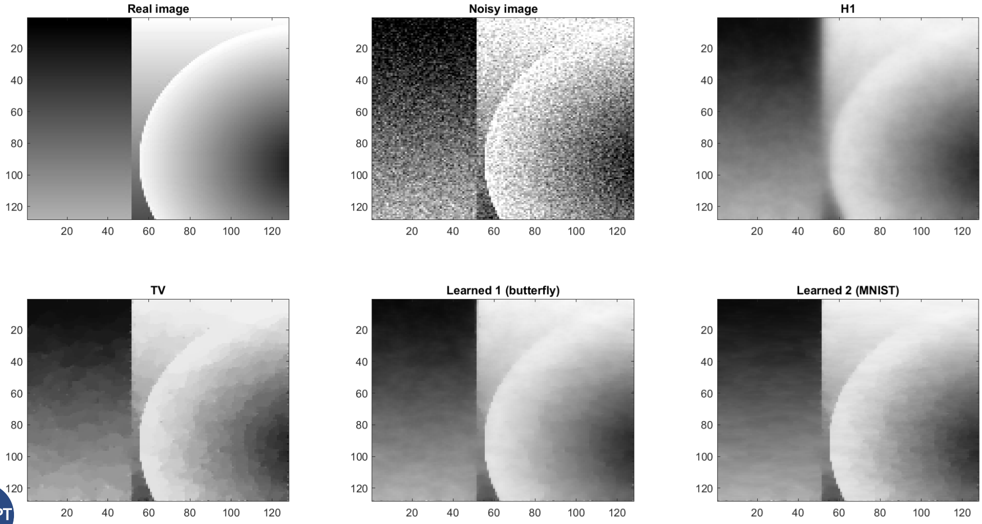


Figure: Classic regularizers compared to ours with noise level $\tau = 30 \approx 10\%$.

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

Conclusions

Contributions:

- ⇝ Supervised learning framework for learning firmly nonexpansive operators,
- ⇝ PA approximations to construct fne operators in practice,
- ⇝ Application to image denoising,

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

Conclusions

Contributions:

- ⇝ Supervised learning framework for learning firmly nonexpansive operators,
- ⇝ PA approximations to construct fne operators in practice,
- ⇝ Application to image denoising,

Challenge:

⇝ Computational intractability of

$$\operatorname*{arg\,min}_{N \in \mathcal{N}} \frac{1}{n} \sum_{i=1}^{n} \|N(\bar{x}_i) - \bar{u}_i\|^2.$$

Is there a better way to solve it in practice?

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

¡Muchas gracias!

(**contact:** jonathanchirinosrodriguez@gmail.com)

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