Data driven regularization by projection

Andrea Aspri

RICAM
JOHANN-RADON-INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Joint work with Y. Korolev and O. Scherzer

Joint meeting Fudan University and RICAM

Shanghai & Linz - 10, June 2020

Motivation

Given $Au = y$ and y_δ noisy measurements s.t. $\|y - y_\delta\| \leq \delta$.

Goal:

Given measurements y_δ, *reconstruct* the unknown quantity u.

Assume we have

$$\{u^i, y^i\}_{i=1,\cdots,n} \text{ s.t. } Au^i = y^i$$

Questions

- How can we use these pairs in the reconstruction process?
- How can we use these pairs when A is *not explicitly known*?

Learning an operator: is it possible?

Main goal: develop stable algorithms for finding u such that $Au = y$

- without *explicit knowledge of* A,

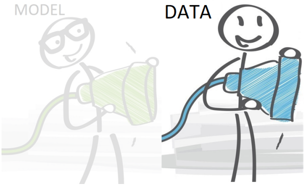
but having only

- *training pairs*:

$$\{u^i, y^i\}_{i=1,\cdots,n} \text{ s.t. } Au^i = y^i$$

- *noisy measurements* y_δ, s.t.

$$\|y - y_\delta\| \leq \delta$$



Question

Is there a regularization method capable of learning a linear operator?

Spoiler: Yes, there is...

Learning an operator: is it possible?

Main goal: develop stable algorithms for finding u such that $Au = y$

- without *explicit knowledge of A*,

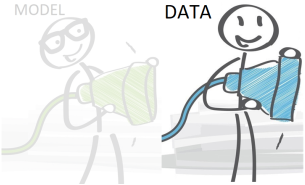
but having only

- *training pairs*:

$$\{u^i, y^i\}_{i=1,\cdots,n} \text{ s.t. } Au^i = y^i$$

- *noisy measurements* y_δ, s.t.

$$\|y - y_\delta\| \leq \delta$$



Question

Is there a regularization method capable of learning a linear operator?

Spoiler: Yes, there is...

Terminology & Notation

u^i are "training images"

$\mathcal{U}_n := \mathrm{Span}\{u^i\}_{i=1,\cdots,n}$

$P_{\mathcal{U}_n}$ orthogonal projection onto $\mathcal{U}_n$

y^i are "training data"

$\mathcal{Y}_n := \mathrm{Span}\{y^i\}_{i=1,\cdots,n}$

$P_{\mathcal{Y}_n}$ orthogonal projection onto $\mathcal{Y}_n$

$u^\dagger$ solution to $Au = y$.

Main assumptions

1. Operator
   - $A : \mathcal{U} \to \mathcal{Y}$, with $\mathcal{U}$, $\mathcal{Y}$ Hilbert spaces.
   - A is bounded, linear and injective (but A^{-1} is unbounded).
2. Data
   - **Linear independence** of $\{u^i\}_{i=1,\cdots,n}$, $\forall n \in \mathbb{N}$.
   - **Uniform boundedness:** $\exists C_u > 0$ s.t. $\|u^i\| \leq C_u$, $\forall i \in \mathbb{N}$.
   - **Sequentiality:** training pairs are nested, i.e.

$$\{u^i, y^i\}_{i=1,\cdots,n+1} = \{u^i, y^i\}_{i=1,\cdots,n} \cup \{u^{n+1}, y^{n+1}\},$$
$$\text{hence}$$
$$\mathcal{U}_n \subset \mathcal{U}_{n+1} \ \text{ and } \ \mathcal{Y}_n \subset \mathcal{Y}_{n+1}, \forall n \in \mathbb{N}.$$

   - **Density:** training images spaces are dense in $\mathcal{U}$, i.e., $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{U}_n} = \mathcal{U}$.

Consequences:

- Training data y^i are *linearly independent* and *uniformly bounded* as well;
- Training data spaces are dense in $\overline{\mathcal{R}(A)}$.

Main assumptions

1. Operator
   - $A : \mathcal{U} \to \mathcal{Y}$, with $\mathcal{U}$, $\mathcal{Y}$ Hilbert spaces.
   - A is bounded, linear and injective (but A^{-1} is unbounded).
2. Data
   - **Linear independence** of $\{u^i\}_{i=1,\cdots,n}$, $\forall n \in \mathbb{N}$.
   - **Uniform boundedness:** $\exists C_u > 0$ s.t. $\|u^i\| \leq C_u$, $\forall i \in \mathbb{N}$.
   - **Sequentiality:** training pairs are nested, i.e.

$$\{u^i, y^i\}_{i=1,\cdots,n+1} = \{u^i, y^i\}_{i=1,\cdots,n} \cup \{u^{n+1}, y^{n+1}\},$$
$$\text{hence}$$
$$\mathcal{U}_n \subset \mathcal{U}_{n+1} \ \text{ and } \ \mathcal{Y}_n \subset \mathcal{Y}_{n+1}, \forall n \in \mathbb{N}.$$

   - **Density:** training images spaces are dense in $\mathcal{U}$, i.e., $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{U}_n} = \mathcal{U}$.

Consequences:

- Training data y^i are *linearly independent* and *uniformly bounded* as well;
- Training data spaces are dense in $\overline{\mathcal{R}(A)}$.

Main assumptions

1. Operator
   - $A : \mathcal{U} \to \mathcal{Y}$, with $\mathcal{U}$, $\mathcal{Y}$ Hilbert spaces.
   - A is bounded, linear and injective (but A^{-1} is unbounded).
2. Data
   - **Linear independence** of $\{u^i\}_{i=1,\cdots,n}$, $\forall n \in \mathbb{N}$.
   - **Uniform boundedness:** $\exists C_u > 0$ s.t. $\|u^i\| \leq C_u$, $\forall i \in \mathbb{N}$.
   - **Sequentiality:** training pairs are nested, i.e.

$$\{u^i, y^i\}_{i=1,\cdots,n+1} = \{u^i, y^i\}_{i=1,\cdots,n} \cup \{u^{n+1}, y^{n+1}\},$$

hence

$$\mathcal{U}_n \subset \mathcal{U}_{n+1} \text{ and } \mathcal{Y}_n \subset \mathcal{Y}_{n+1}, \forall n \in \mathbb{N}.$$

   - **Density:** training images spaces are dense in $\mathcal{U}$, i.e., $\overline{\bigcup_{n\in\mathbb{N}} \mathcal{U}_n} = \mathcal{U}$.

Consequences:

- Training data y^i are *linearly independent* and *uniformly bounded* as well;
- Training data spaces are dense in $\overline{\mathcal{R}(A)}$.

Regularization by projection

Approximate $u^\dagger$ using the Minimum Norm Solution (MNS) of finite dimensional problems

Least-Squares Proj.

(1) $AP_n u = y$

- $P_n =$ orthogonal projection onto a finite dimensional space of $\mathcal{U}$.

Dual Least-Squares Proj.

(2) $Q_n A u = Q_n y$

- Q_n orthogonal projection onto a finite dimensional space of $\mathcal{Y}$.

Our idea to use training pairs

- Choose $P_n = P_{\mathcal{U}_n}$.

It can be proven

$$\text{MNS}: \quad u_n^{\mathcal{U}} = A^{-1} P_{\mathcal{Y}_n} y$$

In general $u_n^{\mathcal{U}} \nrightarrow u^\dagger$.

- Choose $Q_n = P_{\mathcal{Y}_n}$

It can be proven

$$\text{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^*\mathcal{Y}_n} u_n^{\mathcal{U}}$$

In this case $u_n^{\mathcal{Y}} \rightarrow u^\dagger$.

Engl, H. W. and Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Springer (1996).

Seidman T. I., Nonconvergence results for the application of least-squares estimation to ill-posed problems, J. Opt. Th. Appl. (1980).

Regularization by projection

Approximate $u^\dagger$ using the Minimum Norm Solution (MNS) of finite dimensional problems

Least-Squares Proj.	*Dual Least-Squares Proj.*
(1) $AP_n u = y$	(2) $Q_n A u = Q_n y$
• $P_n =$ orthogonal projection onto a finite dimensional space of $\mathcal{U}$.	• Q_n orthogonal projection onto a finite dimensional space of $\mathcal{Y}$.

Our idea to use training pairs

• Choose $P_n = P_{\mathcal{U}_n}$.	• Choose $Q_n = P_{\mathcal{Y}_n}$
It can be proven	It can be proven
$\mathrm{MNS}: \quad u_n^{\mathcal{U}} = A^{-1} P_{\mathcal{Y}_n} y$	$\mathrm{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^* \mathcal{Y}_n} u_n^{\mathcal{U}}$
In general $u_n^{\mathcal{U}} \nrightarrow u^\dagger$.	In this case $u_n^{\mathcal{Y}} \rightarrow u^\dagger$.

Engl, H. W. and Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Springer (1996).

Seidman T. I., Nonconvergence results for the application of least-squares estimation to ill-posed problems, J. Opt. Th. Appl. (1980).

Regularization by projection

Approximate $u^\dagger$ using the Minimum Norm Solution (MNS) of finite dimensional problems

Least-Squares Proj.

$$(1) \quad AP_n u = y$$

- $P_n =$ orthogonal projection onto a finite dimensional space of $\mathcal{U}$.

Dual Least-Squares Proj.

$$(2) \quad Q_n A u = Q_n y$$

- Q_n orthogonal projection onto a finite dimensional space of $\mathcal{Y}$.

Our idea to use training pairs

- Choose $P_n = P_{\mathcal{U}_n}$.

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{U}} = A^{-1} P_{\mathcal{Y}_n} y$$

In general $u_n^{\mathcal{U}} \nrightarrow u^\dagger$.

- Choose $Q_n = P_{\mathcal{Y}_n}$

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^*\mathcal{Y}_n} u_n^{\mathcal{U}}$$

In this case $u_n^{\mathcal{Y}} \to u^\dagger$.

Engl, H. W. and Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Springer (1996).

Seidman T. I., Nonconvergence results for the application of least-squares estimation to ill-posed problems, J. Opt. Th. Appl. (1980).

Regularization by projection

Approximate $u^\dagger$ using the Minimum Norm Solution (MNS) of finite dimensional problems

Least-Squares Proj.

$$(1) \quad AP_n u = y$$

- $P_n =$ orthogonal projection onto a finite dimensional space of $\mathcal{U}$.

Dual Least-Squares Proj.

$$(2) \quad Q_n A u = Q_n y$$

- Q_n orthogonal projection onto a finite dimensional space of $\mathcal{Y}$.

Our idea to use training pairs

- Choose $P_n = P_{\mathcal{U}_n}$.

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{U}} = A^{-1} P_{\mathcal{Y}_n} y$$

In general $u_n^{\mathcal{U}} \nrightarrow u^\dagger$.

- Choose $Q_n = P_{\mathcal{Y}_n}$

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^* \mathcal{Y}_n} u_n^{\mathcal{U}}$$

In this case $u_n^{\mathcal{Y}} \to u^\dagger$.

Engl, H. W. and Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Springer (1996).

Seidman T. I., Nonconvergence results for the application of least-squares estimation to ill-posed problems, J. Opt. Th. Appl. (1980).

Regularization by projection

Approximate $u^\dagger$ using the Minimum Norm Solution (MNS) of finite dimensional problems

Least-Squares Proj.

$$(1) \quad AP_n u = y$$

- $P_n =$ orthogonal projection onto a finite dimensional space of $\mathcal{U}$.

Dual Least-Squares Proj.

$$(2) \quad Q_n A u = Q_n y$$

- Q_n orthogonal projection onto a finite dimensional space of $\mathcal{Y}$.

Our idea to use training pairs

- Choose $P_n = P_{\mathcal{U}_n}$.

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{U}} = A^{-1} P_{\mathcal{Y}_n} y$$

In general $u_n^{\mathcal{U}} \nrightarrow u^\dagger$.

- Choose $Q_n = P_{\mathcal{Y}_n}$

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^*\mathcal{Y}_n} u_n^{\mathcal{U}}$$

In this case $u_n^{\mathcal{Y}} \rightarrow u^\dagger$.

Engl, H. W. and Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Springer (1996).

Seidman T. I., Nonconvergence results for the application of least-squares estimation to ill-posed problems, J. Opt. Th. Appl. (1980).

Regularization by projection

Approximate $u^\dagger$ using the Minimum Norm Solution (MNS) of finite dimensional problems

Least-Squares Proj.

$$(1) \quad AP_n u = y$$

- $P_n =$ orthogonal projection onto a finite dimensional space of $\mathcal{U}$.

Dual Least-Squares Proj.

$$(2) \quad Q_n A u = Q_n y$$

- Q_n orthogonal projection onto a finite dimensional space of $\mathcal{Y}$.

Our idea to use training pairs

- Choose $P_n = P_{\mathcal{U}_n}$.

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{U}} = A^{-1} P_{\mathcal{Y}_n} y$$

In general $u_n^{\mathcal{U}} \nrightarrow u^\dagger$.

- Choose $Q_n = P_{\mathcal{Y}_n}$

It can be proven

$$\mathrm{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^* \mathcal{Y}_n} u_n^{\mathcal{U}}$$

In this case $u_n^{\mathcal{Y}} \to u^\dagger$.

Engl, H. W. and Hanke, M. and Neubauer, A., Regularization of Inverse Problems, Springer (1996).

Seidman T. I., Nonconvergence results for the application of least-squares estimation to ill-posed problems, J. Opt. Th. Appl. (1980).

Projection onto Image Space (least-squares proj.)

We consider $AP_{\mathcal{U}_n}u = y$ and MNS $u_n^{\mathcal{U}} = A^{-1}P_{\mathcal{Y}_n}y$.

Goal: give an explicit expression of $u_n^{\mathcal{U}}$ in terms of training pairs only and find conditions to guarantee convergence of $u_n^{\mathcal{U}}$ to $u^\dagger$.

Key ingredient: the Gram-Schmidt orthonormalization procedure

$$\{y^i\}_{1=1,\cdots,n} \xrightarrow{G-S} \{\bar{y}^i\}_{i=1,\cdots,n}$$

- $\{\bar{y}^i\}_{i=1,\cdots,n}$ is a orthonormal basis of $\mathcal{Y}_n$.
- We identify $\bar{u}^i$ s.t. $A\bar{u}^i = \bar{y}^i$, for $i = 1, \cdots, n$.

Remark: By Gram-Schmidt it is easy to verify that

$$\bar{u}^i = \frac{u^i - \sum_{k=1}^{i-1}(y^i, \bar{y}^k)\bar{u}^k}{\|y^i - P_{\mathcal{Y}_{i-1}}y^i\|}, \qquad \forall i \in \mathbb{N}$$

Explicit Representation

$$u_n^{\mathcal{U}} = A^{-1}P_{\mathcal{Y}_n}y = \sum_{i=1}^n (y, \bar{y}^i)\bar{u}^i.$$

Projection onto Image Space (least-squares proj.)

We consider $AP_{\mathcal{U}_n}u = y$ and MNS $u_n^{\mathcal{U}} = A^{-1}P_{\mathcal{Y}_n}y$.

Goal: give an explicit expression of $u_n^{\mathcal{U}}$ in terms of training pairs only and find conditions to guarantee convergence of $u_n^{\mathcal{U}}$ to $u^{\dagger}$.

Key ingredient: the Gram-Schmidt orthonormalization procedure

$$\{y^i\}_{1=1,\cdots,n} \xrightarrow{G-S} \{\bar{y}^i\}_{i=1,\cdots,n}$$

- $\{\bar{y}^i\}_{i=1,\cdots,n}$ is a orthonormal basis of $\mathcal{Y}_n$.
- We identify $\bar{u}^i$ s.t. $A\bar{u}^i = \bar{y}^i$, for $i = 1, \cdots, n$.

Remark: By Gram-Schmidt it is easy to verify that

$$\bar{u}^i = \frac{u^i - \sum_{k=1}^{i-1}(y^i, \bar{y}^k)\bar{u}^k}{\|y^i - P_{\mathcal{Y}_{i-1}}y^i\|}, \qquad \forall i \in \mathbb{N}$$

Explicit Representation

$$u_n^{\mathcal{U}} = A^{-1}P_{\mathcal{Y}_n}y = \sum_{i=1}^{n}(y, \bar{y}^i)\bar{u}^i.$$

Question

What about convergence of $u_n^{\mathcal{U}}$?

Need to require some *restrictive* assumptions on training pairs and exact data y.

- Weak convergence iff $\|u_n^{\mathcal{U}}\| \leq C_1,\ C_1 > 0$, for all $n \in \mathbb{N}$.

  To introduce sufficients conditions on the data to guaranteee weak convergence

  we use Gram-Schmidt, $\{u^i\} \xrightarrow{G-S} \{\underline{u}^i\}$ and $A\underline{u}^i = \underline{y}^i$. Then (not optimal) sufficient conditions are

  - $\sum_{j=1}^{+\infty} |(u^\dagger, \underline{u}^j)| < \infty$
  - for all $i \in \mathbb{N}$, given $P_{\mathcal{Y}_n}\underline{y}^i = \sum_{j=1}^n \beta_j^{i,n}\underline{y}^j$, assume that

  $$\sum_{j=1}^n (\beta_j^{i,n})^2 \leq C.$$

- Strong convergence: $\sum_{i=1}^{\infty} |(y, \bar{y}^i)| \|\bar{u}^i\| < \infty$.

Question

What about convergence of $u_n^{\mathcal{U}}$?

Need to require some *restrictive* assumptions on training pairs and exact data y.

- Weak convergence iff $\|u_n^{\mathcal{U}}\| \leq C_1,\ C_1 > 0$, for all $n \in \mathbb{N}$.

  To introduce sufficients conditions on the data to guaranteee weak convergence

  we use Gram-Schmidt, $\{u^i\} \xrightarrow{G-S} \{\underline{u}^i\}$ and $A\underline{u}^i = \underline{y}^i$. Then (not optimal) sufficient conditions are

  - $\sum_{j=1}^{+\infty} |(u^\dagger, \underline{u}^j)| < \infty$
  - for all $i \in \mathbb{N}$, given $P_{\mathcal{Y}_n}\underline{y}^i = \sum_{j=1}^{n} \beta_j^{i,n}\underline{y}^j$, assume that

  $$\sum_{j=1}^{n} (\beta_j^{i,n})^2 \leq C.$$

- Strong convergence: $\sum_{i=1}^{\infty} |(y, \bar{y}^i)| \|\bar{u}^i\| < \infty$.

Noisy data

When we have $y_\delta = y + \Delta$, with $\Delta \notin \mathcal{R}(A)$ and $\|\Delta\| \leq \delta$, we have

$$u_n^{\mathcal{U}} = \sum_{i=1}^{n} (y_\delta, \bar{y}^i)\bar{u}^i$$

We cannot expect convergence (since A^{-1} is unbounded), in fact

$$u^\dagger - u_n^{\mathcal{U}} = A^{-1}(I - P_{\mathcal{Y}_n})Au^\dagger - A^{-1}P_{\mathcal{Y}_n}\Delta.$$

Moral

The number of training pairs plays the role of a regularisation parameter.

Projection onto Data Space (dual least-squares proj.)

$$P_{\mathcal{Y}_n} A u = P_{\mathcal{Y}_n} y \qquad (1)$$

$$\mathrm{MNS}: \quad u_n^{\mathcal{Y}} = P_{A^* \mathcal{Y}_n} u_n^{\mathcal{U}} \qquad (2)$$

Main issue: to implement (2) we need the training pairs

$$(v^i, y^i)_{i=1,\cdots,n}$$

where $v^i = A^* y^i$, $i = 1, \cdots, n$.

Question

What happens when

- restrictive assumptions on the traning pairs are not satisfied?
- data of the adjoint operator are missing?

(A possible) Answer

Use a regularization of the projected problem.

Variational regularization

Let's start again from $AP_{\mathcal{U}_n}u = y$ and $\{u^i, y^i\}_{i=1,\cdots,n}$. We study

Optimization problem

$$\min_{u\in\mathcal{U}} \frac{1}{2}\|AP_{\mathcal{U}_n}u - y_\delta\|^2 + \alpha\mathcal{J}(u)$$

$\alpha=$ regularization parameter, $\mathcal{J}$ a regularization term.

Key ingredient: the Gram-Schmidt orthonormalization procedure

$$\{u^i\}_{1=1,\cdots,n} \xrightarrow{G-S} \{\underline{u}^i\}_{i=1,\cdots,n} \longrightarrow \underline{y}^i := A\underline{u}^i$$

Hence $AP_{\mathcal{U}_n}u = \sum_{i=1}^{n}(u, \underline{u}^i)\underline{y}^i$.

Analysis: existence of a minimiser and convergence rates comes from classical results.

Neubauer, A. and Scherzer, O., Finite-dimensional approximation of Tikhonov regularized solutions of nonlinear ill-posed problems, Num. Func. Anal. Opt. (1990).

Pöschl, C., Resmerita, E. and Scherzer, O., Discretization of variational regularization in Banach spaces, Inv. Prob. (2010).

Variational regularization

Let's start again from $AP_{\mathcal{U}_n}u = y$ and $\{u^i, y^i\}_{i=1,\cdots,n}$. We study

Optimization problem

$$\min_{u\in\mathcal{U}} \frac{1}{2}\|AP_{\mathcal{U}_n}u - y_\delta\|^2 + \alpha\mathcal{J}(u)$$

$\alpha=$ regularization parameter, $\mathcal{J}$ a regularization term.

Key ingredient: the Gram-Schmidt orthonormalization procedure

$$\{u^i\}_{1=1,\cdots,n} \xrightarrow{G-S} \{\underline{u}^i\}_{i=1,\cdots,n} \longrightarrow \underline{y}^i := A\underline{u}^i$$

Hence $AP_{\mathcal{U}_n}u = \sum_{i=1}^n (u, \underline{u}^i)\underline{y}^i$.

Analysis: existence of a minimiser and convergence rates comes from classical results.

Neubauer, A. and Scherzer, O., Finite-dimensional approximation of Tikhonov regularized solutions of nonlinear ill-posed problems, Num. Func. Anal. Opt. (1990).

Pöschl, C., Resmerita, E. and Scherzer, O., Discretization of variational regularization in Banach spaces, Inv. Prob. (2010).

Tests - Setting

Forward operator: Radon transform

$$Ru = y$$

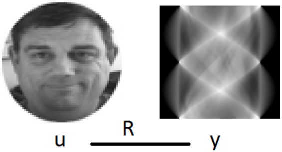


Figure: An example of training pairs.

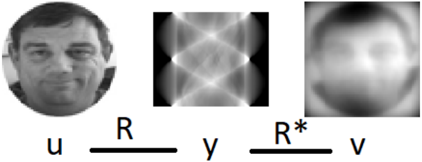


Figure: Examples of triplets used for projection onto data space.

In variational regularization, we choose (for example) $\mathcal{J} = TV$.

Noise 1% (in y_δ)

Original | Least-Sq. | Dual Least-Sq. | Variational Regul.

Figure: Reconstruction using 1000 training pairs and with 1% of noise.

Figure: Reconstruction using 2000 training pairs and with 1% of noise.

Increasing the number of training pairs...

Original | Least-Squares | Dual Least-Squares

Figure: Reconstruction using 2000 training pairs and 1% of noise.

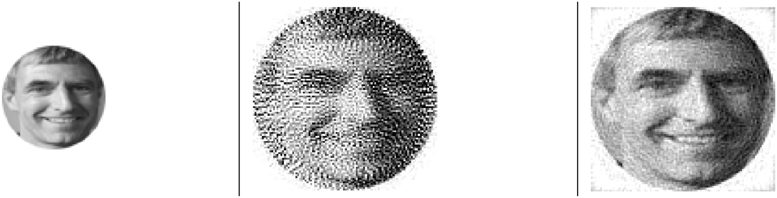


Figure: Reconstruction using 5000 training pairs and 1% of noise.

Increasing the number of training pairs...

Original	Variational Regul. (data-driven)	Variational Regul. (model-based)
		

Figure: Reconstruction using 7000 training pairs and 1% of noise in the data driven variational regularization.

Conclusions

We have seen:

- Model-based regularization theory can be extended to purely data driven setting;
- No numerical access to the forward operator (in the injective case);
- About the size of the traning set:
  - *variational regularization*: having more training data is better;
  - *regularization by projection*: size of training set as regularization parameter $\longrightarrow$ too many training pairs comprise stability;
- Regularization by projection is capable of learning linear (injective) operators.

Thank you for your attention