

Learning in Inverse Problems

Supervised learning

Obvious idea: supervised learning

Use data pairs for input-output related by

$$f^\delta = Ku + \eta$$

Minimize risk with appropriate loss L over some neural network architecture

$$\min_{\theta} \mathbb{E}_{(u, f^\delta)} (L(u, \mathcal{N}_{\theta}(f^\delta))) = \mathbb{E}_{(u, \nu)} (L(u, \mathcal{N}_{\theta}(Ku + \nu)))$$

Issues of supervised learning

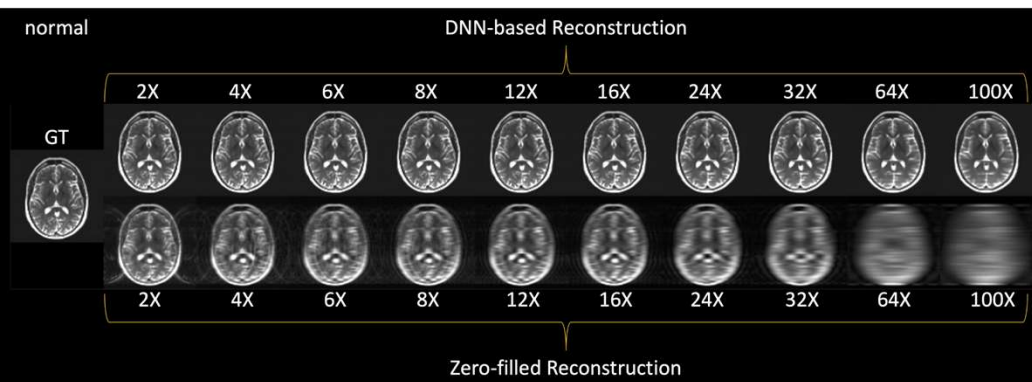
- (Computational) complexity of the inverse problem
- Bad generalization (network for inversion needs huge Lipschitz constant)
- Missing pairs of input-output data

Learning in Image Reconstruction

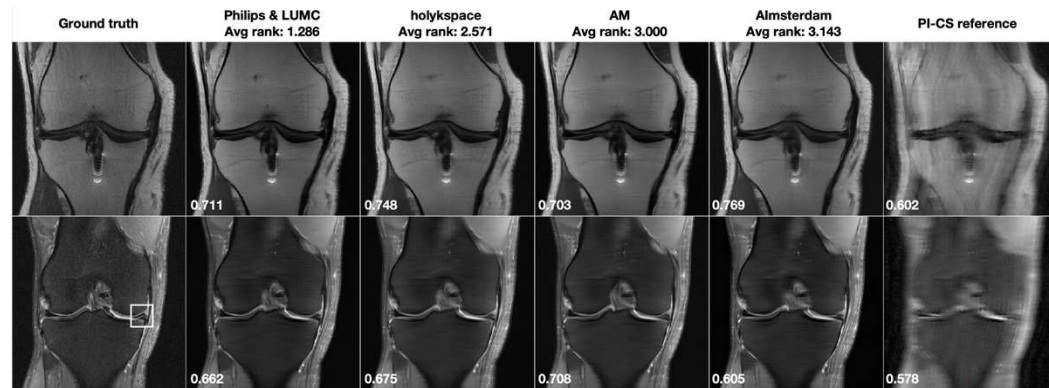
Undersampled MRI

Undersampling in MRI does not suffer from these issues (partly also in CT):

- Lower complexity, since forward operator just Fourier transform, low noise
- Isometry property of Fourier transform leads to low Lipschitz constant of inverse
- Data pairs from existing fully sampled measurements and reconstructions



Radmanesh Radiology AI 2022



Knoll MRM 2019 / 2020 (fast MRI challenge)

Learning in Image Reconstruction

Undersampled MRI

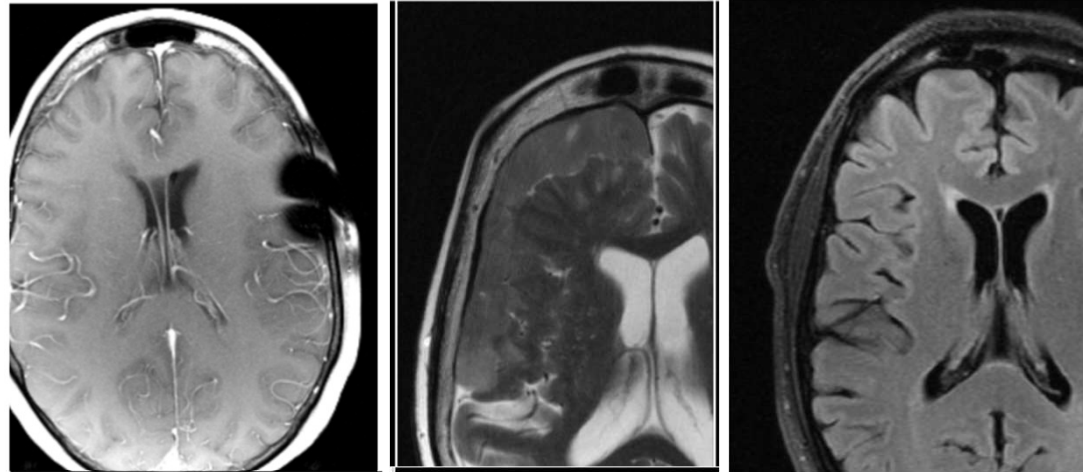
Majority of results convincing

But possible hallucinations
on few data sets

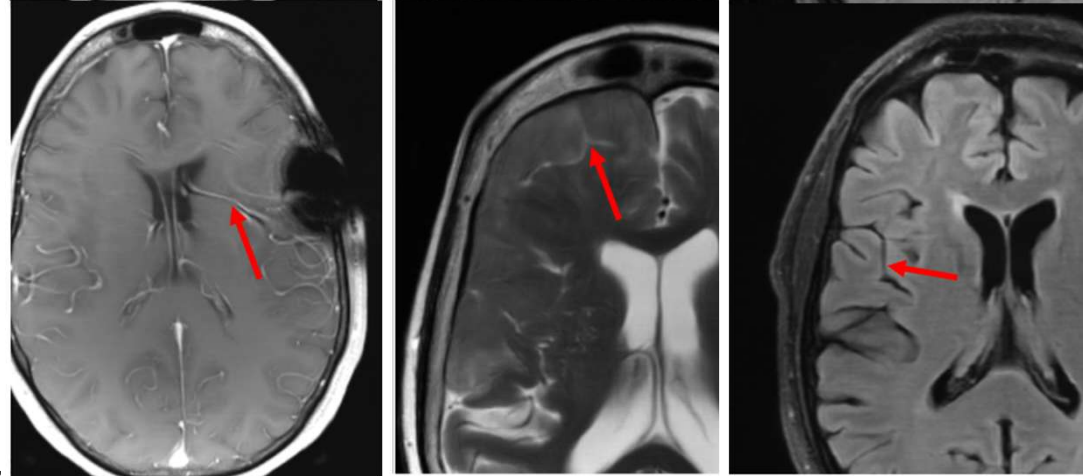
Not recognizable by
experienced radiologists

(courtesy Florian Knoll, Erlangen)

Ground truth



Reconstruction



Muckley TMI 2021

Further issues in supervised learning

„Semi-Supervised learning“

Paradigm: still solve

$$\hat{u} \in \arg \min_u \left(F(Ku, f^\delta) + \alpha J(u) \right)$$

but with regularizer J (and possibly regularization parameter) learned from a database of images
(and possibly unrelated noisy data)

Bayesian interpretation: directly learn prior, form posterior with forward model

Examples

- Adversarial regularizations
- Plug and play priors: trained by denoising on images solely
- Score-based diffusion models: transform prior into Gaussian, construct biased Langevin sampling to go back to approximate sampling of posterior

Further issues in supervised learning

Adversarial regularizers

Example: adversarial learning [Lunz-Öktem-Schönlieb 18]

Given favourable images $\{u_i\}_{i=1}^n$ and unfavourable ones $\{v_k\}_{k=1}^m$
minimize (with respect to parameters)

$$\frac{1}{n} \sum_{i=1}^n J(u_i) - \frac{1}{m} \sum_{k=1}^m J(v_k) + \lambda \mathbb{E}[(\|\nabla J\| - 1)_+^2]$$

Learned regularization method is itself a random variable in terms of training data.

As n and m tend to infinity and under assumption of i.i.d. sampling from appropriate distributions expect convergence to minimizer of deterministic population risk

$$\mathbb{E}_u(J) - \mathbb{E}_v(J) + \lambda \mathbb{E}[(\|\nabla J\| - 1)_+^2]$$

Detailed properties of regularizer and subsequent solutions of inverse problem remain unclear

So far, functionals learned based on data sets, but independent of inverse problem (forward operator K).

Unclear if training data could even be solution of inverse problem

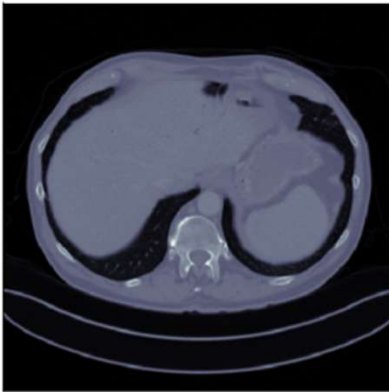
Learned Regularizers

Adversarial regularization with source condition

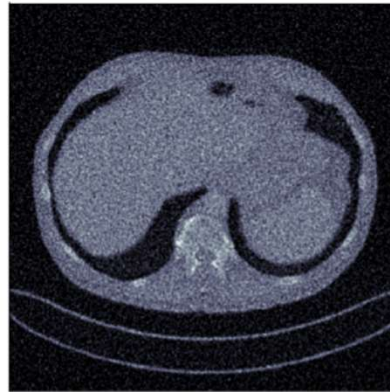
Augment with penalty that ensures training data satisfy source condition [mb-Mukherjee-Schönlieb, NeurIPS Workshop 2021]

$$\frac{1}{n} \sum_{i=1}^n \|(K^*)^{-1} \partial_u J(u_i)\|^2 \quad \mathbb{E}_u(\|(K^*)^{-1} \partial J(u)\|^2)$$

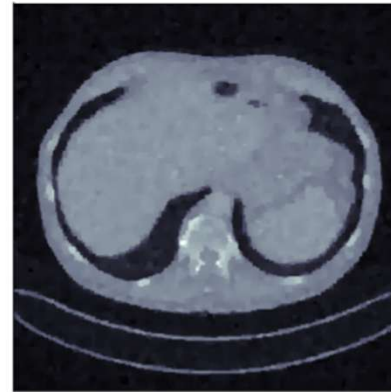
Undersampled and noisy CT reconstruction (Mayo Clinic Low Dose dataset)



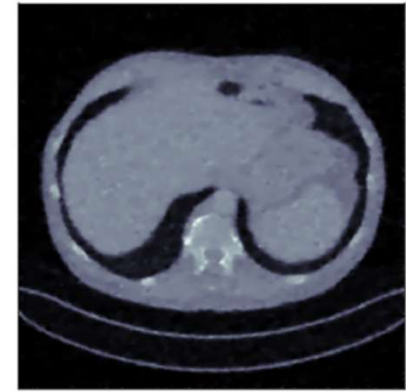
(a) ground-truth



(b) FBP: 21.19, 0.22



(c) TV: 29.85, 0.79



(h) ACR-SC: 30.93, 0.85

Learned Regularizers

Adversarial regularization with source condition

Best case comparison:

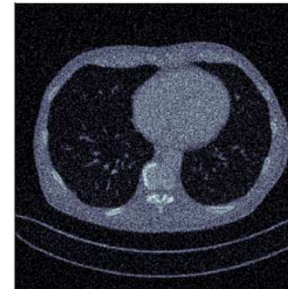
Supervised learning and methods
with less constraints superior

However, more interpretability
and robustness with source condition
constraint

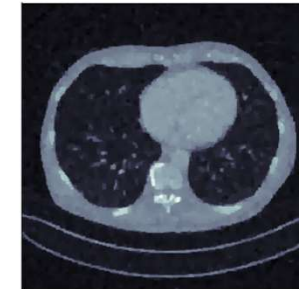
method	PSNR (dB)	SSIM	# param.	reconstruction time (ms)
FBP	21.28 ± 0.13	0.20 ± 0.02	1	37.0 ± 4.6
TV	30.31 ± 0.52	0.78 ± 0.01	1	28371.4 ± 1281.5
<i>Supervised methods</i>				
U-Net	34.50 ± 0.65	0.90 ± 0.01	7215233	44.4 ± 12.5
LPD	35.69 ± 0.60	0.91 ± 0.01	1138720	279.8 ± 12.8
<i>Unsupervised methods</i>				
AR	33.84 ± 0.63	0.86 ± 0.01	19338465	22567.1 ± 309.7
ACR	31.55 ± 0.54	0.85 ± 0.01	606610	109952.4 ± 497.8
ACR-SC	31.28 ± 0.50	0.84 ± 0.01	590928	105232.1 ± 378.5



(i) ground-truth



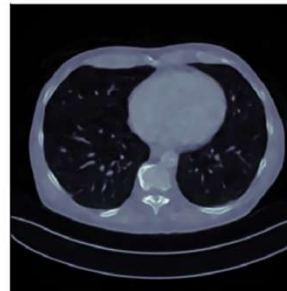
(j) FBP: 21.59, 0.24



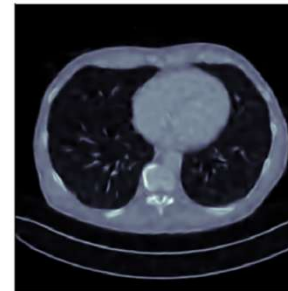
(k) TV: 29.16, 0.77



(l) U-net: 32.69, 0.87



(m) LPD: 34.05, 0.89



(n) AR: 32.14, 0.84



(o) ACR: 30.14, 0.83



(p) ACR-SC: 29.88, 0.82

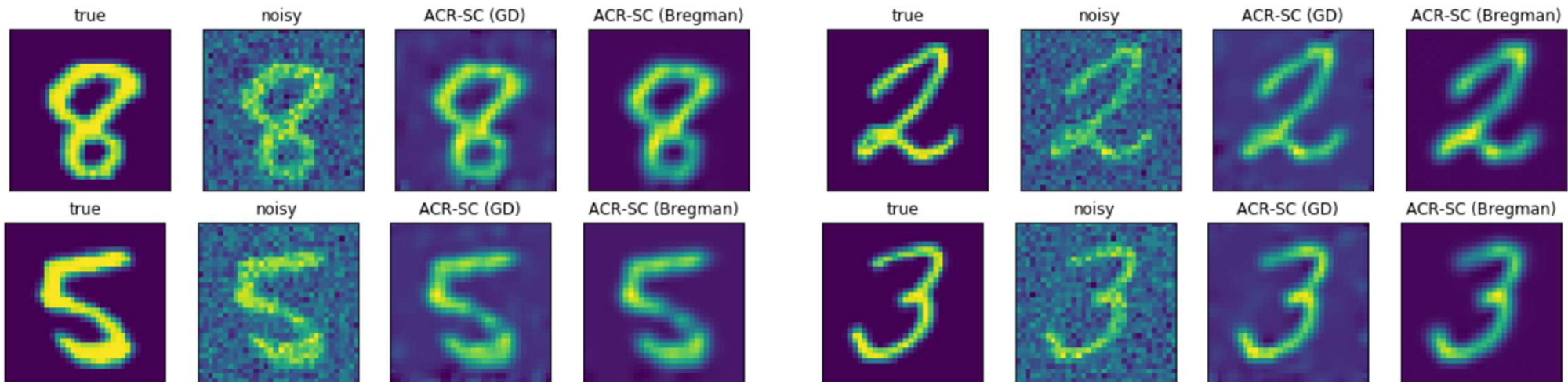
Learned Regularizers

Adversarial regularization with source condition

Additional advantage: interpretable method allows to use superior approaches developed for variational models

Example: Bregman Iteration for Bias Correction, iterative recentering of prior. **Mean SSIM improvement > 10 %**
[Bregman 1967] [Hestenes 1969, Powell 1969] [Osher-mb-Goldfarb-Xu-Yin 2005]

$$u^{k+1} \in \arg \min_u F(Ku, f) + \frac{1}{\tau} \left(J(u) - J(u^k) - \langle p^k, u - u^k \rangle \right)$$
$$p^{k+1} = p^k + \tau K^* \partial F(Ku^{k+1}, f)$$



Score-based Diffusion Models

Generating images

Popular diffusion models can be formulated in a similar form. Given an empirical distribution of samples, evolve towards Gaussian (β used to accelerate)

$$dx = -\beta(t)X dt + \sqrt{2\beta(t)}dW$$
$$\partial_t \mu = \beta(t) \nabla \cdot (\mu x + \nabla \mu)$$

Learn score (or potential) as a neural network from (Song et al 2020)

$$\int_0^T \lambda(t) \int |s_\theta(x, t) - \nabla \log \mu|^2 \mu(dx) dt$$

To generate novel samples use Gaussian samples as starting values and solve the backward SDE

$$d\tilde{X} = \beta(T - \tilde{t})\tilde{X} d\tilde{t} + \sqrt{2\beta(T - \tilde{t})} d\tilde{W} - 2\beta(T - \tilde{t})s_\theta(\tilde{X}, T - \tilde{t})$$

Single Particle Imaging with Lenses

Uncertain Ptychography

Single particle imaging takes projections of crystals or proteins at random orientation

Effectively Radon transform with random unknown orientation

Novel development: lenses for imaging in FELs, similar to ptychography (Bajt, Chapman et al 2022-2025)

Application to Single Particle Imaging: unknown orientation and position

Model problem: Blind ptychography.

Measure absolute value of Fourier transform of object multiplied with mask at unknown location

Welker-Kuger-Roith-mb-Gerkmann-Chapman 2025

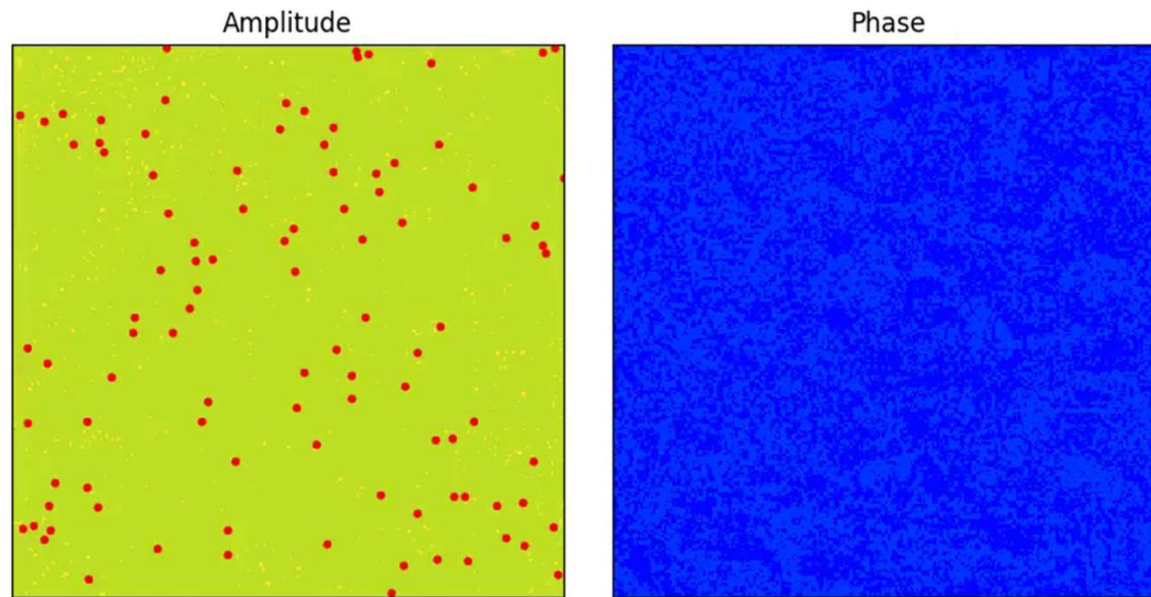
Blind Ptychography

Diffusion

To deal with high uncertainty we use a diffusion model as regularizer

Backward pass towards likelihood of measurement with variational inference

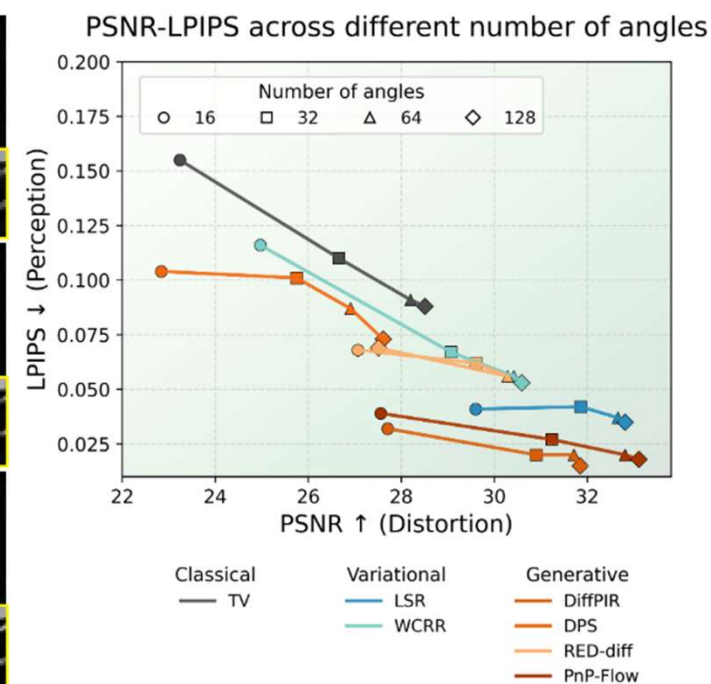
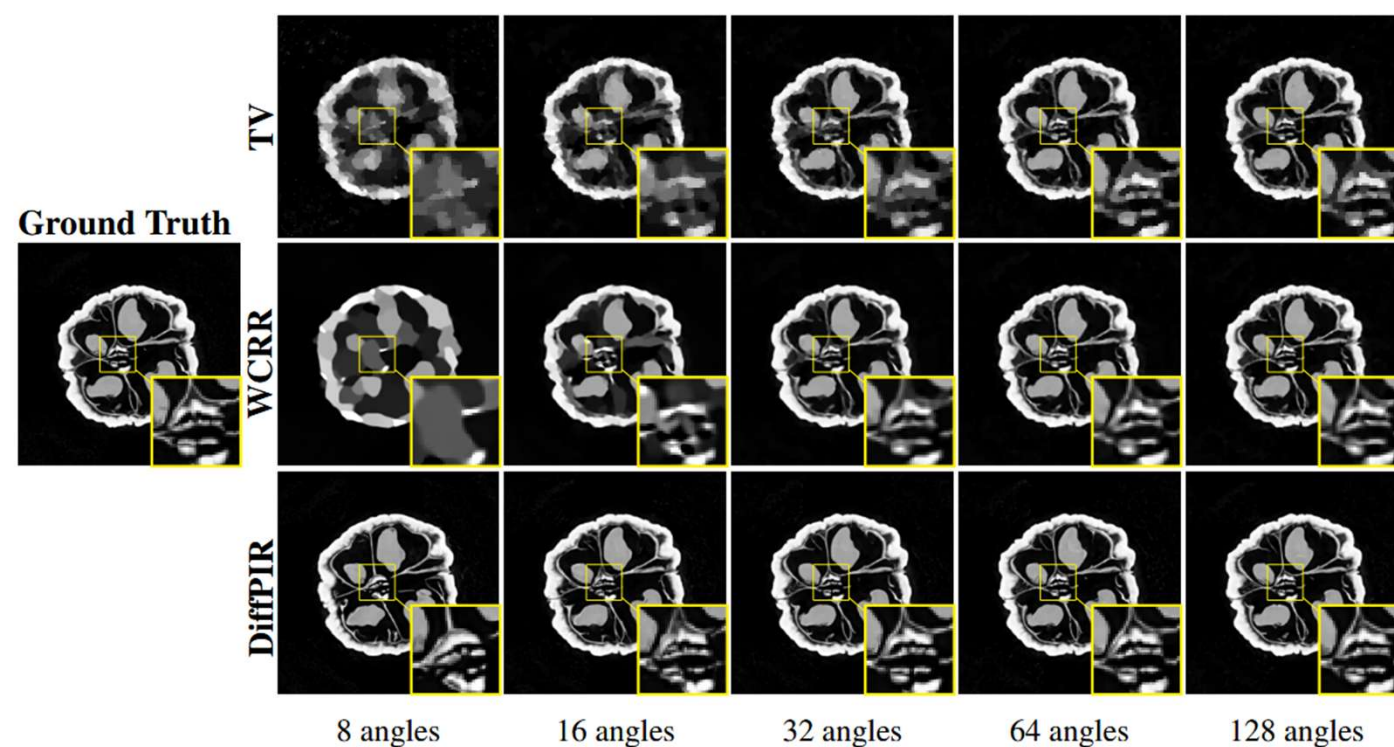
Iterated in EM-type fashion with optimization of mask locations



A Stability Benchmark of Generative Regularizers for Inverse Problems

Denker et al. (preprint, arXiv:2605.10076)

- Evaluation of stability and reliability of generative priors for scientific imaging (computed tomography)
- Benchmark generative priors against modern optimization/variational approaches



When information are missing, variational methods (e.g., WCRR) tend to „wash out“ the unknown regions, whereas generative priors (e.g., DiffPIR) hallucinate realistic-looking structures.

Learning in Image Reconstruction

State of the Art

Common approach:

- Use appropriate neural network for data at fixed resolution
- Use appropriate, often synthetic data set to train
- Display results and compare with reconstruction method that do not use any training data
- Find out that learning surprisingly leads results that look better

Few approaches to provide theoretical insights, often in finite dimension or with assumptions that make the original image reconstruction problem well-posed

2019 | OriginalPaper | Buchkapitel

Deep Learning for Trivial Inverse Problems

verfasst von : Peter Maass

Erschienen in: Compressed Sensing and Its Applications

Verlag: Springer International Publishing

Deep neural networks can stably solve high-dimensional, noisy,
non-linear inverse problems

Andrés Felipe Lerma Pineda*

Philipp Christian Petersen†

Learning the optimal Tikhonov regularizer for inverse problems

Giovanni S. Alberti¹, Ernesto De Vito¹, Matti Lassas², Luca Ratti¹, Matteo Santacesaria¹

OPEN ACCESS

IOP Publishing

Inverse Problems 38 (2022) 115005 (21pp)

Inverse Problems

<https://doi.org/10.1088/1361-1462/ac9001>

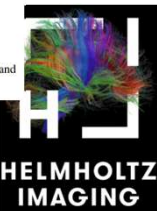
Regularization theory of the analytic deep prior approach

Clemens Arndt*

IEEE SP MAGAZINE SPECIAL ISSUE ON PHYSICS-DRIVEN MACHINE LEARNING FOR COMPUTATIONAL IMAGING

Learned reconstruction methods with convergence guarantees

Subhadip Mukherjee^{*1}, Andreas Hauptmann^{*2,3}, Ozan Öktem⁴, Marcelo Pereyra⁵, and
Carola-Bibiane Schönlieb¹



Page 13

Learning in Image Reconstruction

Open issues

- How do learned methods behave in the infinite-dimensional limit ?
- Do learned methods provide regularization with respect to data noise ? (Guarantees in certain metrics)
- How do typical solutions of a learned regularization method look like ? (Smoothness, bias, ..)
- What is the impact of the specific training approach
- Generalization aspect: do we obtain a convergent regularization method with high probability when trained on finite data ?
-

Learning in Image Reconstruction

A way to theory

As usual in deep learning we face the question whether we can proof anything or give at least a theoretical insight

Possible answer consider simplified model

Here: learning a spectral regularizer [Bauermeister-mb-Möller 2021][Kabri-Auras-Riccio-Benning-Möller-mb 23]

Use singular value decomposition of forward operator K

$$K u = \sum_{n=1}^{\infty} \sigma_n \langle u, u_n \rangle v_n$$

Setup: noise independent of u , unbiased

$$f^\delta = K u + \eta$$

$$\mathbb{E}(\eta) = 0$$

Learning in Image Reconstruction

A way to theory

Supervised learning: from pairs of ground truth and noisy data learn spectral regularization function g in

$$R(f^\delta; g^\delta) = \sum_n g^\delta(\sigma_n) \langle f^\delta, v_n \rangle u_n$$
$$\bar{g} = \arg \min_g \mathbb{E}_{u, \nu} (\|u - R(f^\delta; g)\|^2) \quad g_n = g(\sigma_n)$$

Semi-supervised learning: look for functional of the form (with adversarial approaches as before)

$$J(u) = \sum_{n=1}^{\infty} \lambda_n \langle u, u_n \rangle^2$$

Learning in Image Reconstruction

Four learning approaches

Supervised learning (reduce to learn)

$$\arg \min \mathbb{E}_{u, \nu} (\|u - R(f^\delta; g)\|^2)$$

Bayesian model with Gauss assumption (trivial covariance, MAP = CM)

$$\Sigma_u = \text{diag}(\Pi_n), \quad \Sigma_\eta = \text{diag}(\Delta_n)$$

$$\hat{u} = \arg \min \|Ku - v\|_{\Sigma_\eta^{-1}}^2 + \|u\|_{\Sigma_u^{-1}}^2$$

Adversarial learning

$$\mathbb{E}_u(J(u)) - \mathbb{E}_{u, \eta}(J(u + K^{-1}\nu)) + \frac{\beta}{2} \mathbb{E}_{u, \eta, t}(\|\partial J(u_t)\|^2)$$

Adversarial learning with source condition

$$\mathbb{E}_u(J(u)) - \mathbb{E}_{u, \eta}(J(u + K^{-1}\nu)) + \frac{\beta}{2} \mathbb{E}_u(\|(K^*)^{-1} \partial J(u)\|^2)$$

Learning in Image Reconstruction

Four learning approaches

Models allow for explicit computation of minimizers and comparison

Assume for simplicity zero mean on data set

$$\mathbb{E}_u(u) = 0$$

Notation: power and noise in frequency

$$\Pi_n = \mathbb{E}_u(\langle u, u_n \rangle^2)$$

$$\Delta_n = \mathbb{E}_\eta(\langle \eta, v_n \rangle^2)$$

Learning in Image Reconstruction

1. Supervised learning

Minimizing expected loss with respect to parameters

$$\begin{aligned}\mathbb{E}_{u, \nu} [\|u - R(f^\delta; g)\|^2] &= \mathbb{E} \left[\sum_n (1 - \sigma_n g_n)^2 \langle u, u_n \rangle^2 + g_n^2 \langle \nu, v_n \rangle^2 - 2 \langle u, u_n \rangle \langle \nu, v_n \rangle \right] \\ &= \sum_n (1 - \sigma_n g_n)^2 \Pi_n + g_n^2 \Delta_n,\end{aligned}$$

Leads to

$$\bar{g}_n = \frac{\sigma_n \Pi_n}{\sigma_n^2 \Pi_n + \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\Delta_n}{\Pi_n}}$$

Training with Noise
is Equivalent to Tikhonov Regularization

Christopher M. Bishop
Current address:
Microsoft Research,
7 J J Thomson Avenue,
Cambridge, CB3 0FB, U.K.
cmishop@microsoft.com

<http://research.microsoft.com/~cmbishop>

Published in *Neural Computation* 7 No. 1 (1995) 108-116.

Abstract

It is well known that the addition of noise to the input data of a neural network during training can, in some circumstances, lead to significant improvements in generalization performance. Previous work has shown that such training with noise is equivalent to a form of regularization in which an extra term is added to the error function. However, the regularization term, which involves second derivatives of the error function, is not bounded below, and so can lead to difficulties if used directly in a learning algorithm based on error minimization. In this paper we show that, for the purposes of network training, the regularization term can be reduced to a positive definite form which involves only first derivatives of the network mapping. For a sum-of-squares error function, the regularization term belongs to the class of generalized Tikhonov regularizers. Direct minimization of the regularized error function provides a practical alternative to training with noise.



Learning in Image Reconstruction

2. Bayes

Due to diagonal choice

$$\Sigma_u = \text{diag}(\Pi_n), \quad \Sigma_\eta = \text{diag}(\Delta_n)$$

Leads again to effective regularization

$$\hat{u} = \sum_{n=1}^{\infty} g_n \langle v, v_n \rangle u_n$$
$$\bar{g}_n = \frac{\sigma_n \Pi_n}{\sigma_n^2 \Pi_n + \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\Delta_n}{\Pi_n}}$$

Learning in Image Reconstruction

3./4. Adversarial

For the adversarial methods compute

$$\hat{u} = \arg \min \|Ku - f^\delta\|^2 + \alpha J(u)$$

With diagonal J we have

$$\begin{aligned} & \mathbb{E}_u(J(u)) - \mathbb{E}_{u,\eta}(J(u + K^{-1}\nu)) \\ &= \mathbb{E}_u\left(\sum_n \lambda_n \langle u, u_n \rangle^2\right) - \mathbb{E}_{u,\eta}\left(\sum_n \lambda_n \langle u, u_n \rangle^2 + \sum_n \frac{\lambda_n}{\sigma_n^2} \langle \nu, v_n \rangle^2\right) \\ &= -\sum_n \frac{\lambda_n}{\sigma_n^2} \Delta_n \end{aligned}$$

Learning in Image Reconstruction

Effective spectral regularization

Effective regularization

$$\hat{u} = \sum_{n=1}^{\infty} g_n \langle v, v_n \rangle u_n$$

Supervised / Bayes

$$g_n = \frac{\sigma_n \Pi_n}{\sigma_n^2 \Pi_n + \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\Delta_n}{\Pi_n}}$$

Adversarial

$$g_n = \frac{\sigma_n (2\Pi_n \sigma_n^2 + \Delta_n)}{\sigma_n^2 (2\Pi_n \sigma_n^2 + \Delta_n) + \frac{3\alpha}{\beta} \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\alpha}{\beta} \frac{3\Delta_n}{2\Pi_n \sigma_n^2 + \Delta_n}}$$

Adversarial with source condition

$$g_n = \frac{\sigma_n \Pi_n}{\sigma_n^2 \Pi_n + \frac{\alpha}{\beta} \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\alpha}{\beta} \frac{\Delta_n}{\Pi_n}}$$

Learning in Image Reconstruction

Effective spectral regularization

Effective regularization of high frequencies under white noise

$$\Delta_n \sim \delta \quad \Pi_n \rightarrow 0$$

Supervised / Bayes / Adversarial SC

$$\frac{\sigma_n}{\sigma_n^2 + \frac{\Delta_n}{\Pi_n}} \sim \sigma_n \frac{\Pi_n}{\delta}$$

Adversarial learning becomes like standard Tikhonov

$$\frac{\sigma_n}{\sigma_n^2 + \frac{\alpha}{\beta} \frac{3\Delta_n}{2\Pi_n\sigma_n^2 + \Delta_n}} \sim \sigma_n \frac{\beta}{3\alpha}$$

Learning in Image Reconstruction

Do we reconstruct training data ?

Range condition

$$u \in \mathcal{R}(R(\cdot, \bar{g}))$$

If and only if

$$\sum_n \frac{\Delta_n^2}{\Pi_n^2 \sigma_n^2} \langle u, u_n \rangle^2 < \infty$$

Compare source condition

$$\sum_n \frac{1}{\Pi_n^2} \langle w, u_n \rangle^2 < \infty, \quad u = K^* w$$

Learning in Image Reconstruction

Expected smoothness of reconstructions

Define

$$\tilde{\Pi}_n = \mathbb{E}_{u,\nu}(\langle R(Ku + \nu, \bar{g}), u_n \rangle^2)$$

Can be computed efficiently

$$\tilde{\Pi}_n = \frac{\sigma_n^2 \Pi_n}{\sigma_n^2 \Pi_n + \Delta_n} \Pi_n = \frac{1}{1 + \frac{\Delta_n}{\sigma_n^2 \Pi_n}} \Pi_n$$

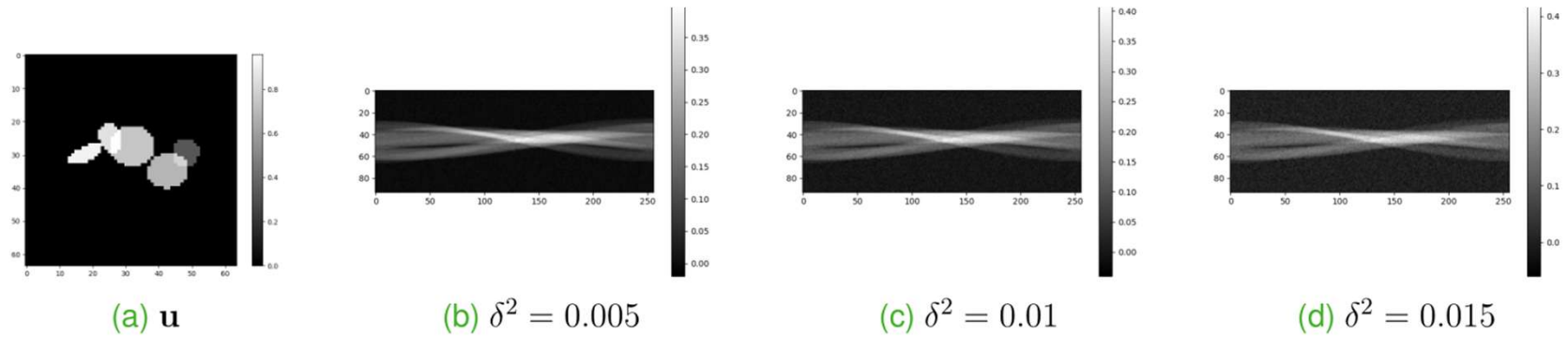
Note: reconstructed solutions are always smoother than clean training data (oversmoothed)

Explanation from regularization applied to rough noise

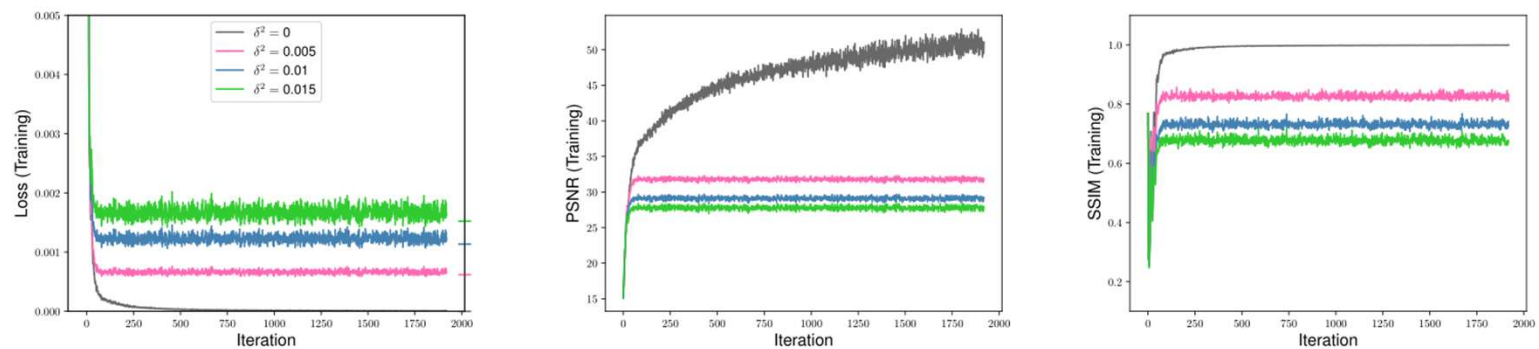
Learning in Image Reconstruction

Spectral regularization in tomography

Training data for different noise level



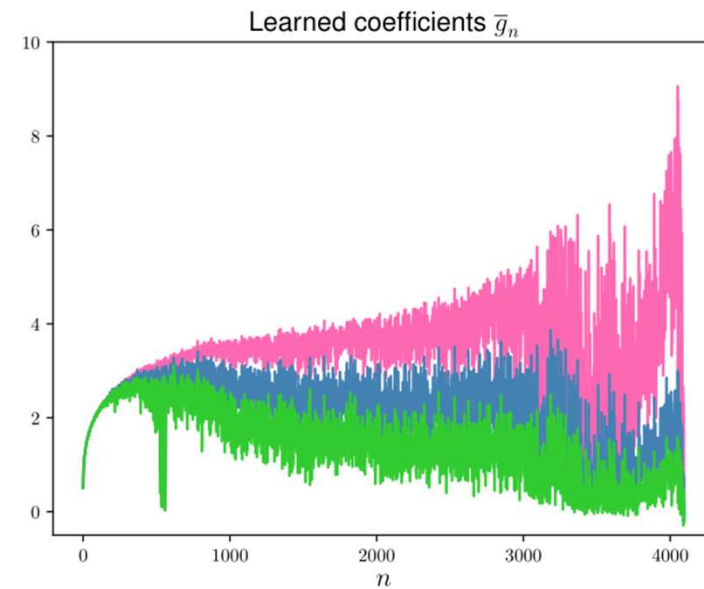
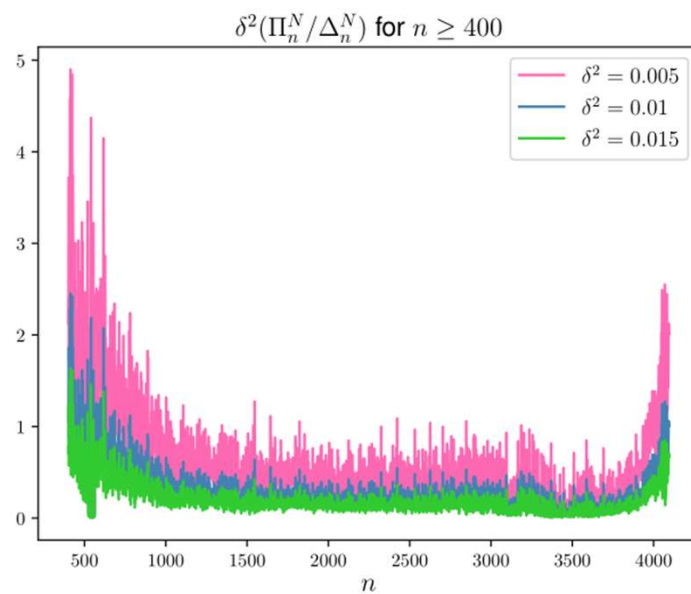
Training statistics



Learning in Image Reconstruction

Spectral regularization in tomography

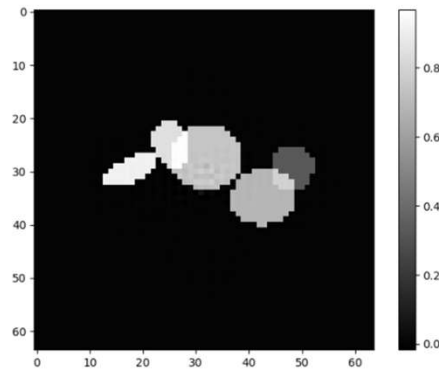
Oversmoothing in computational results



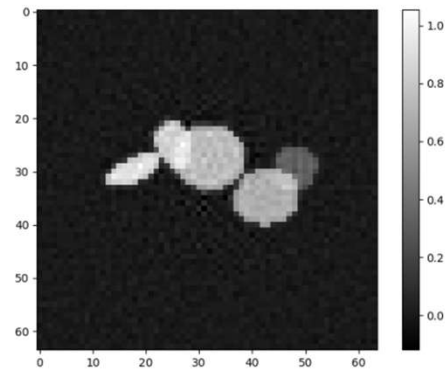
Learning in Image Reconstruction

Spectral regularization in tomography

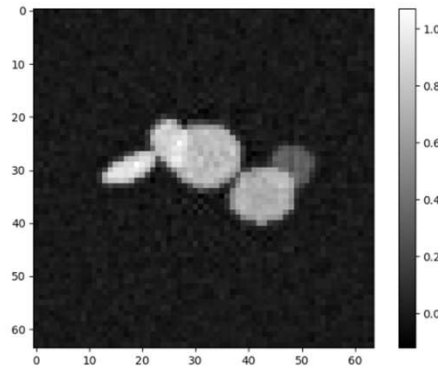
Typical reconstructions



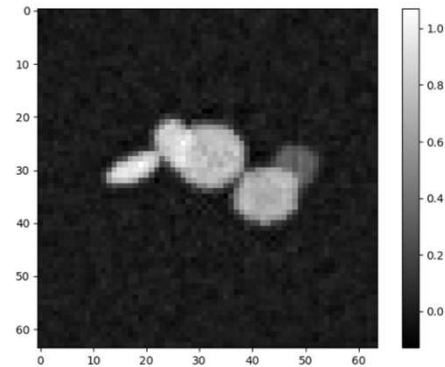
(a) $\delta^2 = 0$



(b) $\delta^2 = 0.005$



(c) $\delta^2 = 0.01$



(d) $\delta^2 = 0.015$

Convergence of learned regularization

Consider a sequence of learned regularizations with respect to noise level²⁹

$$\bar{g}_n^\delta = \frac{\sigma_n \Pi_n}{\sigma_n^2 \Pi_n + \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\Delta_n}{\Pi_n}}$$

Assumptions:

(1) Variance of noise is bounded by δ^2 , more precisely

$$\delta^2 = \max_n \Delta_n.$$

(2) Clean data is smoother than noise, more precisely there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that

$$\Delta_n \geq c \delta^2 \Pi_n$$

for all $n \geq n_0$.

(3) The sequence of Π_n is summable, i.e.,

$$\sum_n \Pi_n < \infty.$$

Convergence of learned regularization

Consider a sequence of learned regularizations with respect to noise level ³⁰

$$\bar{g}_n^\delta = \frac{\sigma_n \Pi_n}{\sigma_n^2 \Pi_n + \Delta_n} = \frac{\sigma_n}{\sigma_n^2 + \frac{\Delta_n}{\Pi_n}}$$

Theorem. *Let Assumptions (1)-(3) hold. Then $R(\cdot; \bar{g}^\delta)$ is continuous and for every $f \in \mathcal{D}(K^\dagger)$*

$$\mathbb{E}_\nu(\|K^\dagger f - R(f^\delta; \bar{g}^\delta)\|^2) \rightarrow 0$$

as $\delta \rightarrow 0$.

Supervised learning

Error

$$\begin{aligned}\|u - R(f^\delta; g)\|^2 &= \left\| \sum_n (\langle u, u_n \rangle - g_n \langle f^\delta, v_n \rangle) u_n \right\|^2 \\ &= \sum_n (\langle u, u_n \rangle - g_n \langle f^\delta, v_n \rangle)^2 \\ &= \sum_n (1 - \sigma_n g_n)^2 \langle u, u_n \rangle^2 + g_n^2 \langle \nu, v_n \rangle^2 - 2 \langle u, u_n \rangle \langle \nu, v_n \rangle\end{aligned}$$

31

Using assumptions on noise

$$\begin{aligned}\mathbb{E}_{u, \nu} [\|u - R(f^\delta; g)\|^2] &= \mathbb{E} \left[\sum_n (1 - \sigma_n g_n)^2 \langle u, u_n \rangle^2 + g_n^2 \langle \nu, v_n \rangle^2 - 2 \langle u, u_n \rangle \langle \nu, v_n \rangle \right] \\ &= \sum_n (1 - \sigma_n g_n)^2 \Pi_n + g_n^2 \Delta_n,\end{aligned}$$

DESY.

$$\Pi_n = \mathbb{E} [\langle u, u_n \rangle^2] \quad \Delta_n = \mathbb{E} [\langle \nu, v_n \rangle^2]$$