The mathematics of image reconstruction

The dialectic of modelling and learning

Martin Burger

DESY Computational Imaging Group / Helmholtz Imaging
Fachbereich Mathematik, Universität Hamburg

HELMHOLTZ

Image Reconstruction

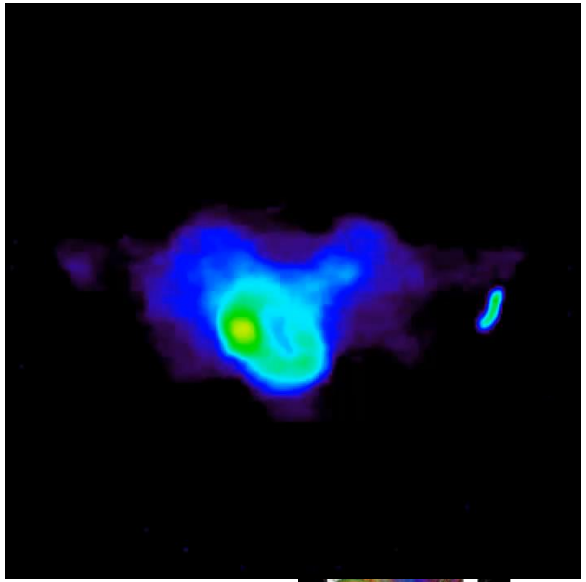
The underestimated part of imaging

Images (Videos) and their manipulation are part of our daily life

First step of image formation often underestimated, although often the enabling part, cf. **CT = Computed** Tomography

Information / quality **loss in** image formation / **reconstruction can hardly be recovered later**

Strong demand on methods for reconstruction and uncertainty quantification in many application fields, from nano to macro



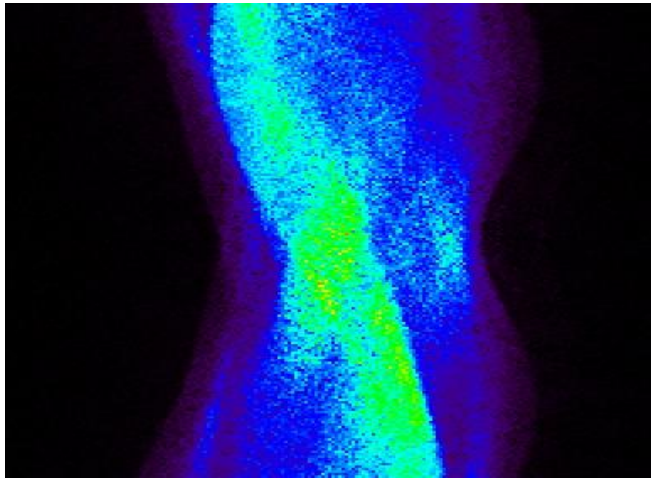
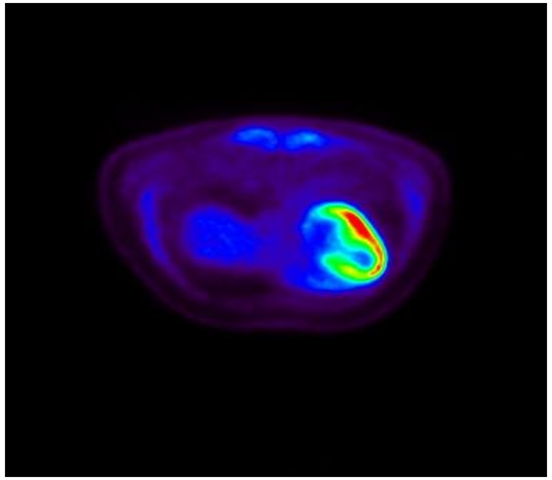
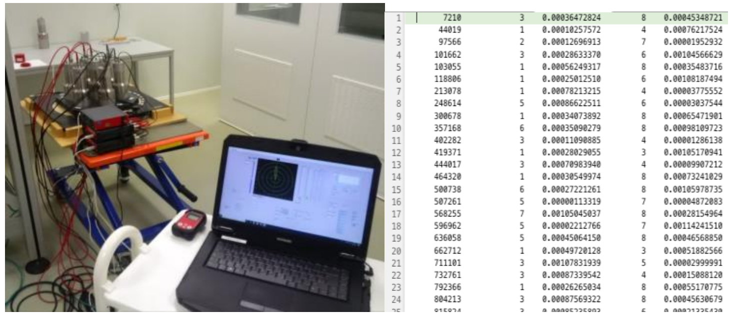
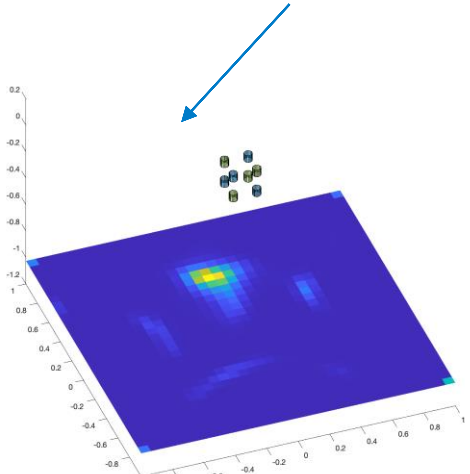
HELMHOLTZ IMAGING

Emission Tomography

Active / Passive

Idea: detect photons emitted e.g. from radioactive decay, with some kind of directional information

- Coincidence based (e.g. PET)
- Collimator based (e.g. SPECT)
- Energy based (Compton effect)
- …

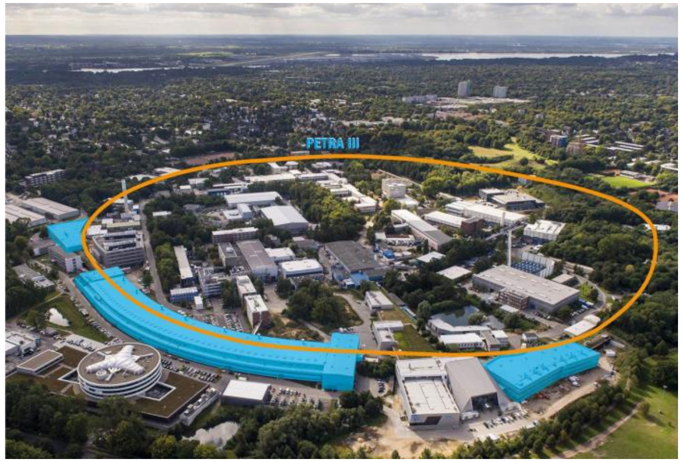





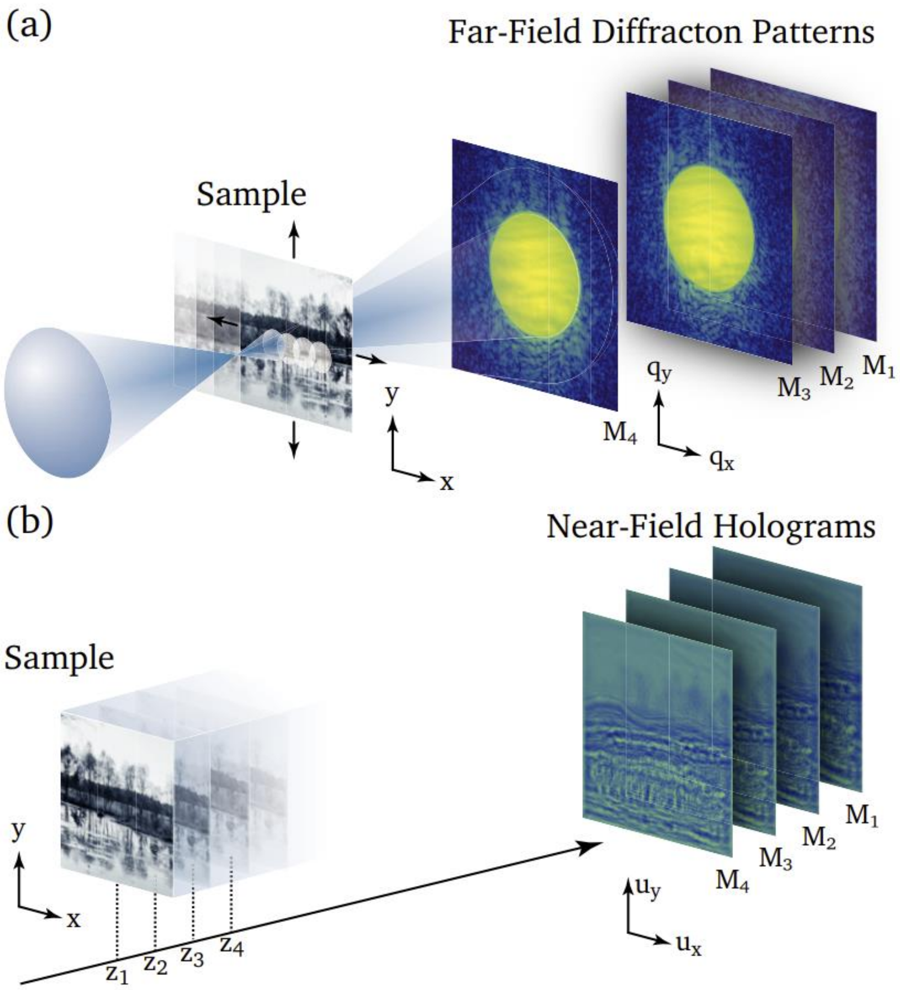
HELMHOLTZ IMAGING

Image reconstruction from synchrotron x-ray sources

Ptychographic / Holographic Tomography

Wittwer et al 2023





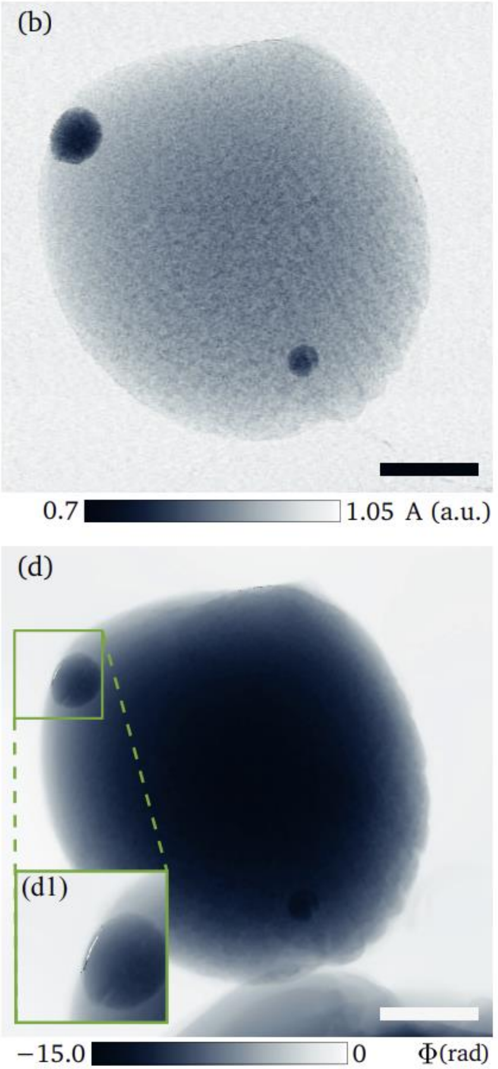
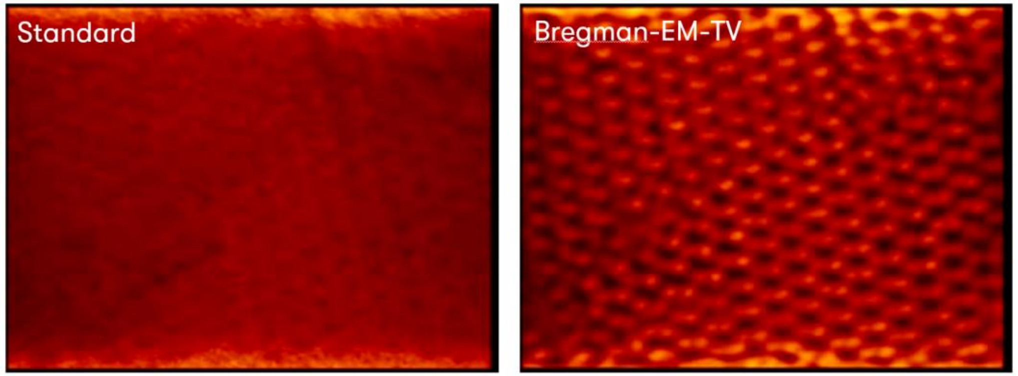
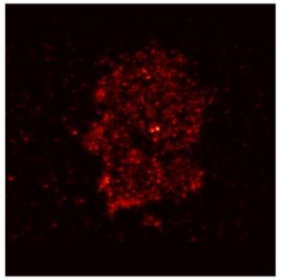


Image reconstruction across scales and planets

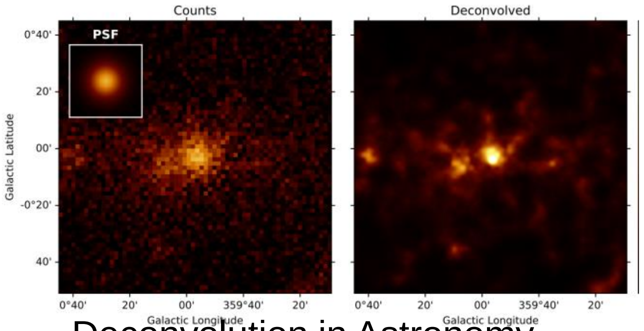
From nano to macro, from intracellular to outer space



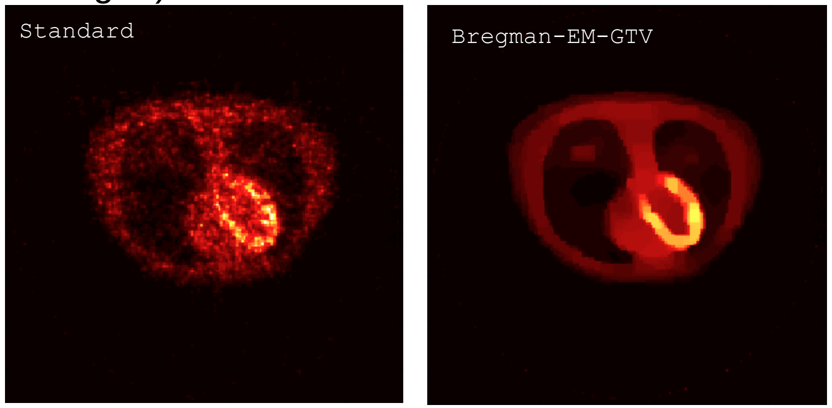
STED Deconvolution of Bead Crystal Structure (with Hell Lab, Göttingen)



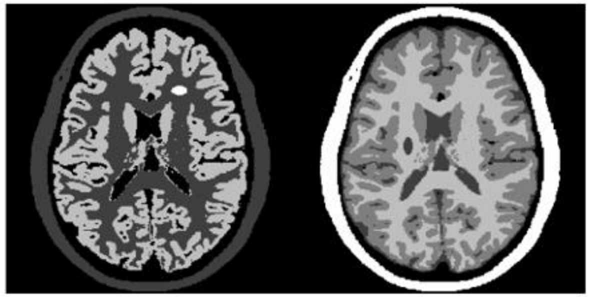
4Pi Deconvolution of Syntaxin PC12 (with Hell Lab, Göttingen)



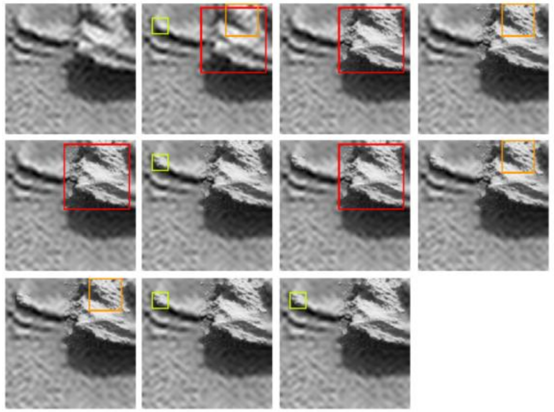
Deconvolution in Astronomy Donath et al 2022



^{18}FDG-PET Reconstruction from short time data (with Nuclear Medicine, Münster)



PET-MR, Rasch-Brinkmann-Burger 2017

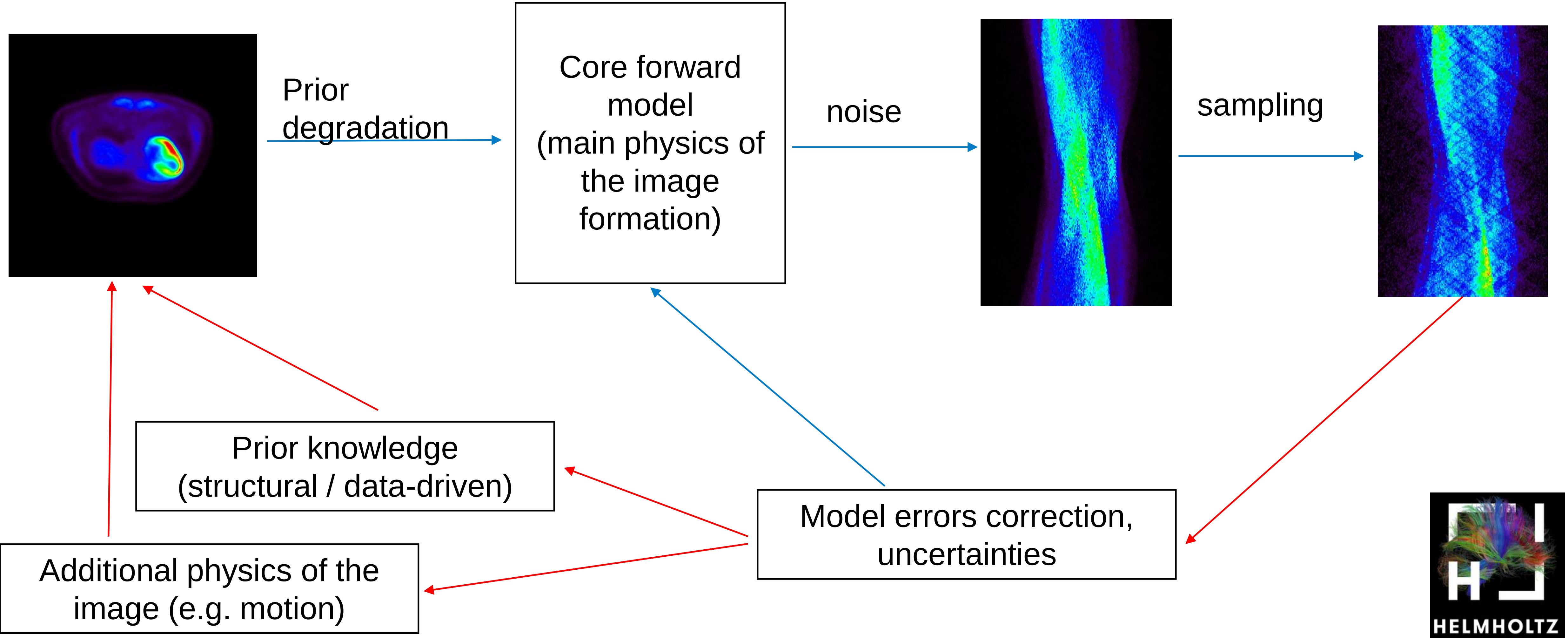


Energy Efficient THZ Imaging on Mars, with DLR Berlin

HELMHOLTZ IMAGING

Modern image reconstruction

Model based view



HELMHOLTZ IMAGING

6

Model based approaches

The classical way of image reconstruction

Formulation as an inverse problem

- Derive physical model of(idealized) forward operator mapping from image to data
- Derive statistical model of noise (e.g. Poisson distribution for photon counts)
- Derive mathematical model of favourable images and structures (e.g. sparsity)
- Possibly add uncertainties

Condensed in Bayesian posterior model

$$\pi(u|f) = \frac{1}{\pi_*(f)} \pi(f|u)\pi_0(u)$$

Likelihood (from u to f) includes forward and noise model**, prior** includes model of favourable images

Model based variational methods

Point estimates

Bayesian **MAP estimate**

$$\hat{u} \in \arg\min_u \left(-\log \pi(f|u) - \log \pi_0(u)\right)$$

Related to **variational regularization method**

$$\hat{u} \in \arg\min_u \Big(F(Ku, f\ \ ) + \alpha J(u) \Big)$$

Forward operator *K*, data fidelity *F*, regularization functional *J*

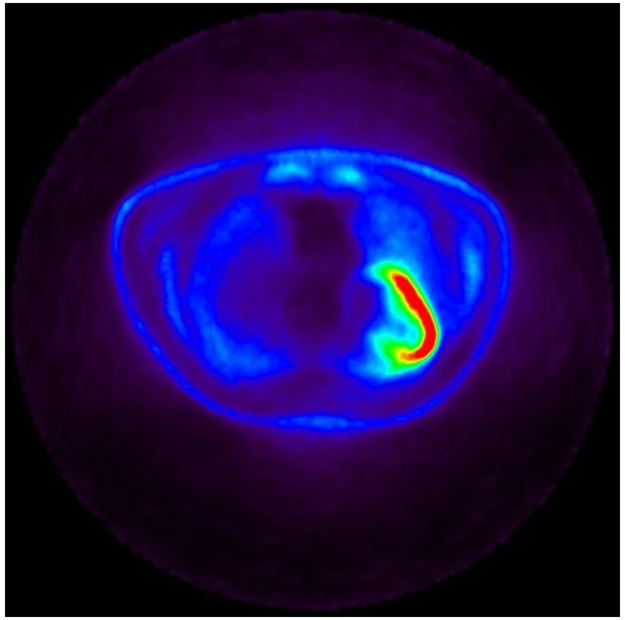
Forward operator: physics (examples: convolution, Radon transform, wave propagation, …)
Data fidelity: stochastics (examples: additive Gaussian noise, Poisson distribution, …)
Regularization: art ? How to translate structural properties into a functional ?

Model based variational methods

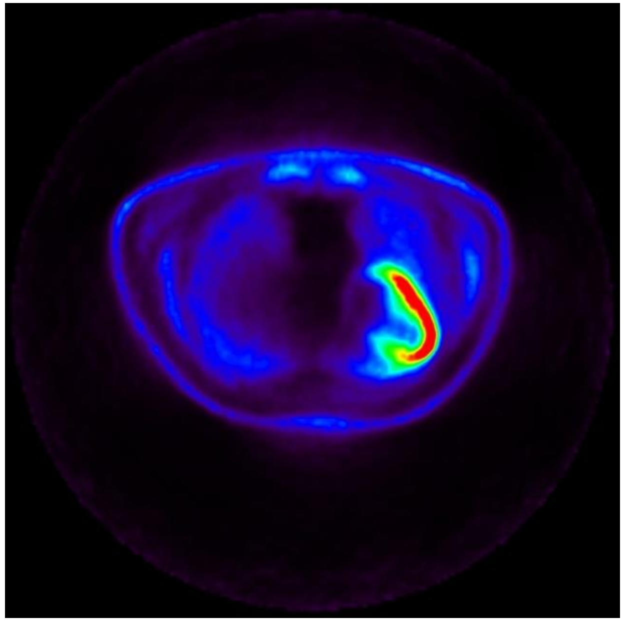
Improving forward models

Example: PET

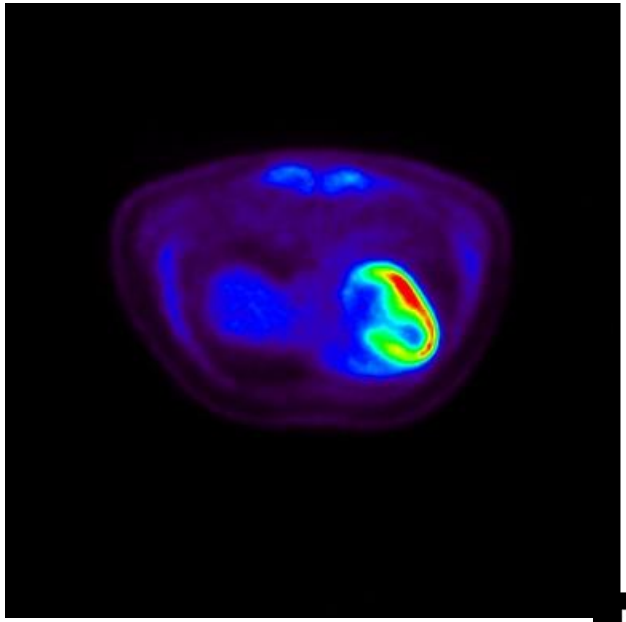
Radon + photon count noise



Radon + photon + scattering



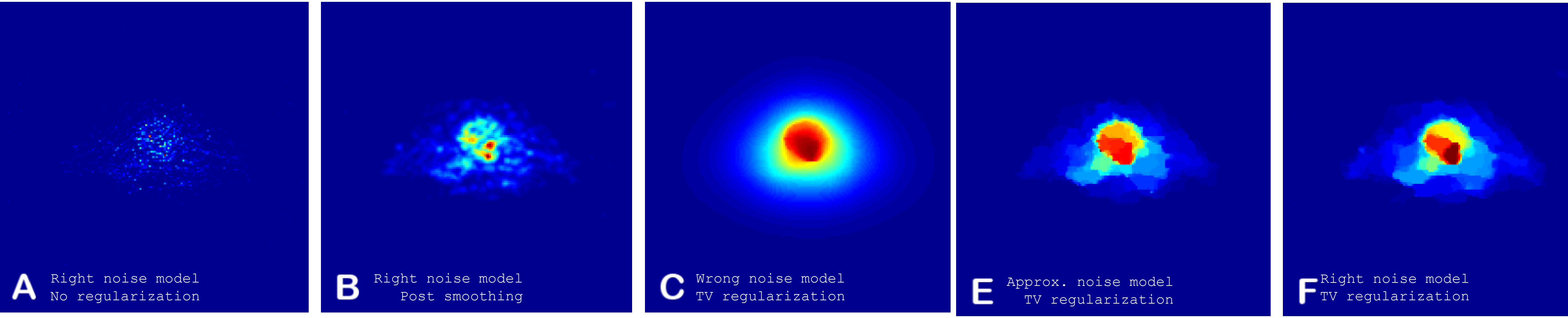
Radon + photon + scattering + attenuation



Model based variational methods

Improving noise models

Example: PET



Cardiac $^{15}H_2O$ PET: Sawatzky, Brune, Müller, Burger 2009

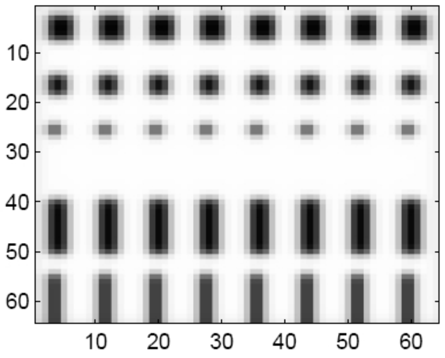
Model based regularizations

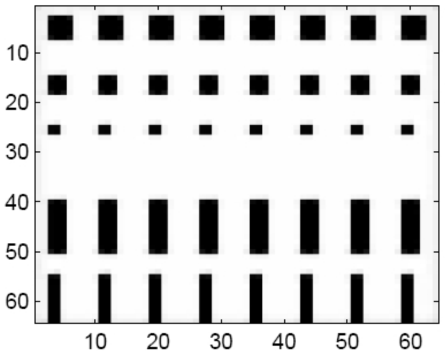
Images with sharp edges

Basic idea from denoising: want to smooth out random noise – local averaging

Simplest idea: Dirichlet energy - quadratic gradient regularization (Gaussian prior)

$$J(u) = \int |\nabla u|^2 \; dx$$





Leads to oversmoothing – no sharp edges

Regularity theory works against us: take $K : L^2 \to Y$

Optimality condition yields $p = -\Delta u = K^* w \in L^2$

Regularity at least $u \in H^2$ does not allow sharp edges

Model based regularizations

Images with sharp edges

Alternative idea: p-Laplacian energy

Similar regularity for p > 1

Limit: total variation

$$TV(u) = |u|_{BV} := \sup_{g \in C_0^\infty(\Omega)^d, g \in \mathcal{C}} \int_\Omega u \nabla \cdot g \, dx$$

$$\mathcal{C} = \{g \in L^\infty(\Omega) \mid |g(x)| \leq 1 \text{ a.e. in } \Omega\}$$

Optimality condition

$$K^* \partial_x F(Ku, f) + \alpha \nabla \cdot g = 0$$

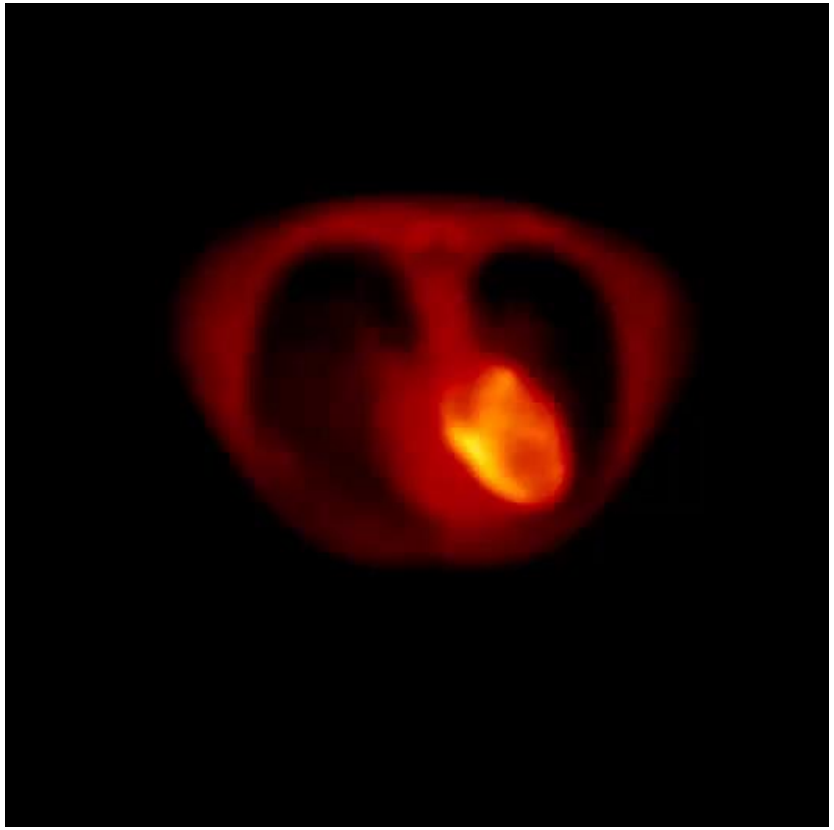
$$g \in \mathcal{C} \qquad \int_\Omega g \cdot dDu = |u|_{BV}$$

Various extensions to cure bias (Bregman iterations) and to avoid staircasing (total generalized variation)

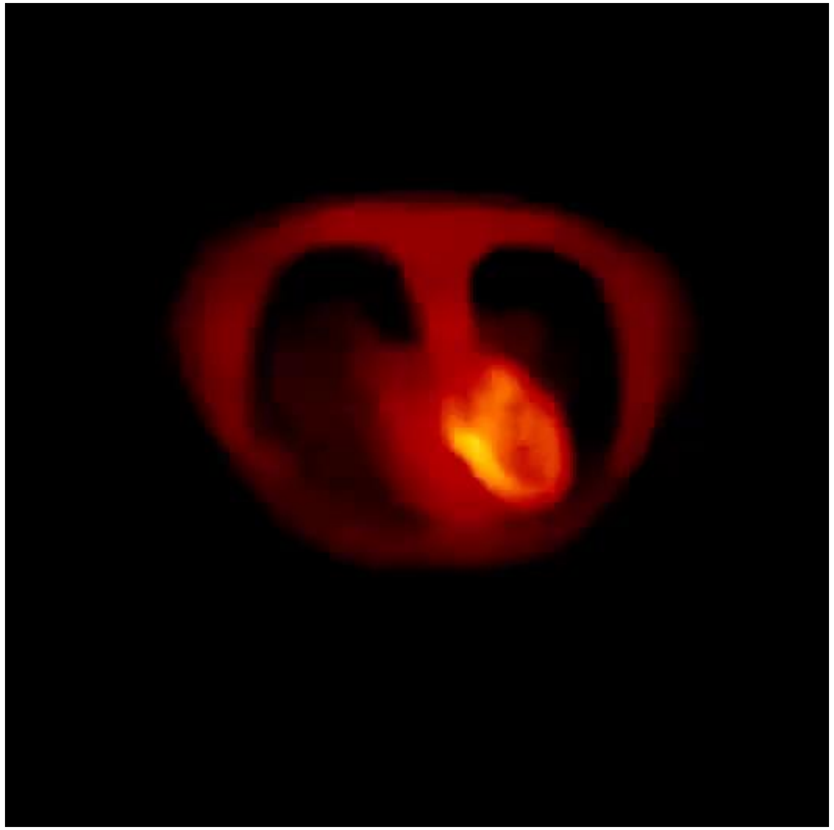
Model based regularizations

Cardiac PET Reconstructios

20 min data, EM reconstruction



5s data, Bregman-TGV regularization



Model based regularizations

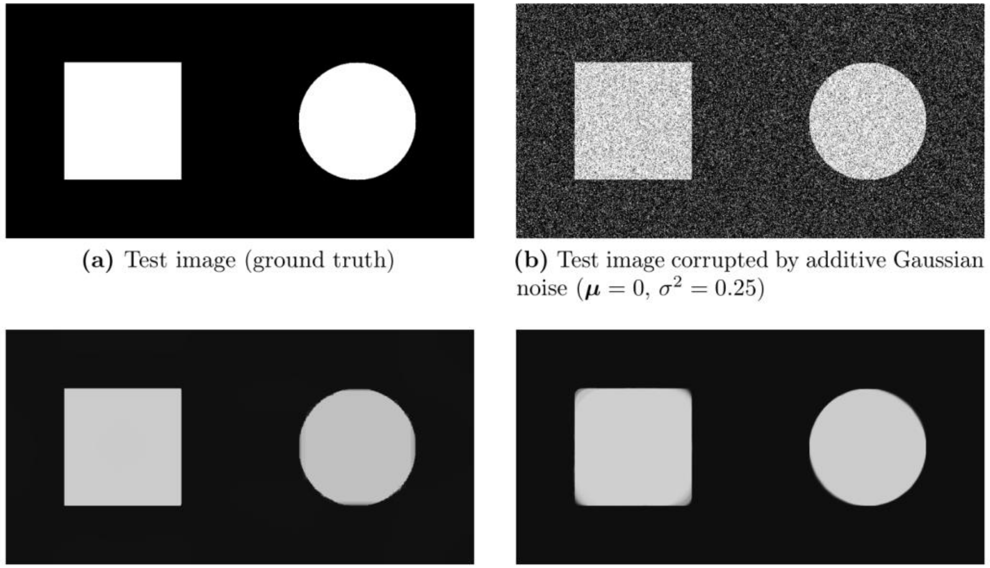
Total variation and related regularization

Optimality (source condition) $$\nabla \cdot g = K^* w$$

g corresponds to (generalized) normal vector field on level lines (surfaces)

Divergence of *g* corresponds to mean curvature

Hence, total variation allows nonsmooth solutions, but smoothes discontinuity sets

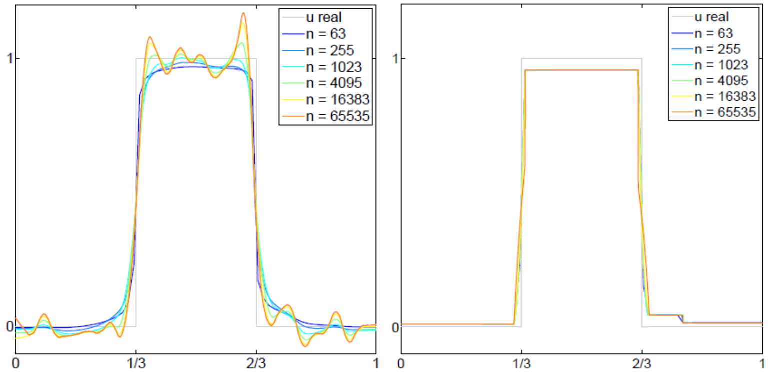


(a) Test image (ground truth)
(b) Test image corrupted by additive Gaussian noise ($\mu = 0$, $\sigma^2 = 0.25$)
(c) Anisotropic TV denoising result ($\alpha = 10$)
(d) Isotropic TV denoising result ($\alpha = 10$)

Problem: modelling very indirect

Prior itself not informative, but only structure of minimizers

Bayesian models for UQ questionable



(a) CM (b) MAP

Model based regularizations

Sparsity regularization

Idea from compressed sensing: choose simple solution (minimal combinations), relax to l1
[Donoho 2006, Candes-Tao 2006]

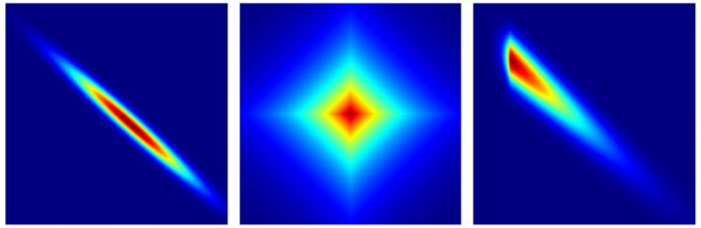
Analysis formulation: for some frame system choose

$$J(u) = \sum_i |\langle u, \phi_i \rangle|$$

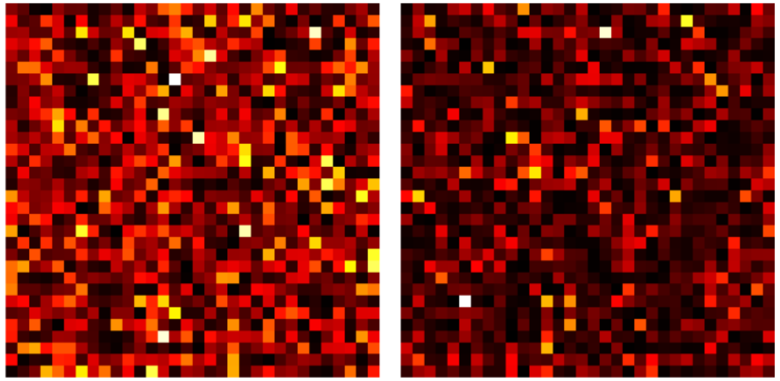
Synthesis formulation (equivalent in case of orthonormal basis)

$$J(u) = \sum_i |c_i| \qquad \text{where } u = \sum_i c_i \phi_i$$

Again similar issue as with TV: based on structure of minimizers
UQ questionable, samples not even sparse



(a) Gaussian likelihood (b) Prior: $p(u) \propto \exp(-\lambda |u|_1)$ (c) Resulting posterior



(a) $p(u) \propto \exp\left(-\frac{1}{2}\|u\|_2^2\right)$ (b) $p(u) \propto \exp\left(-|u|_1\right)$

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

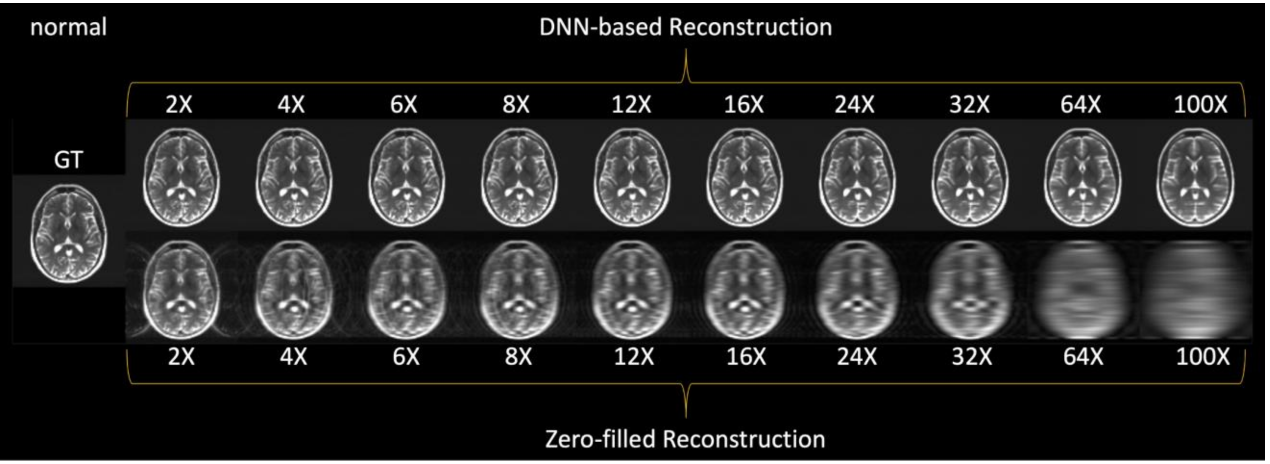
HELMHOLTZ IMAGING

Learning in Image Reconstruction

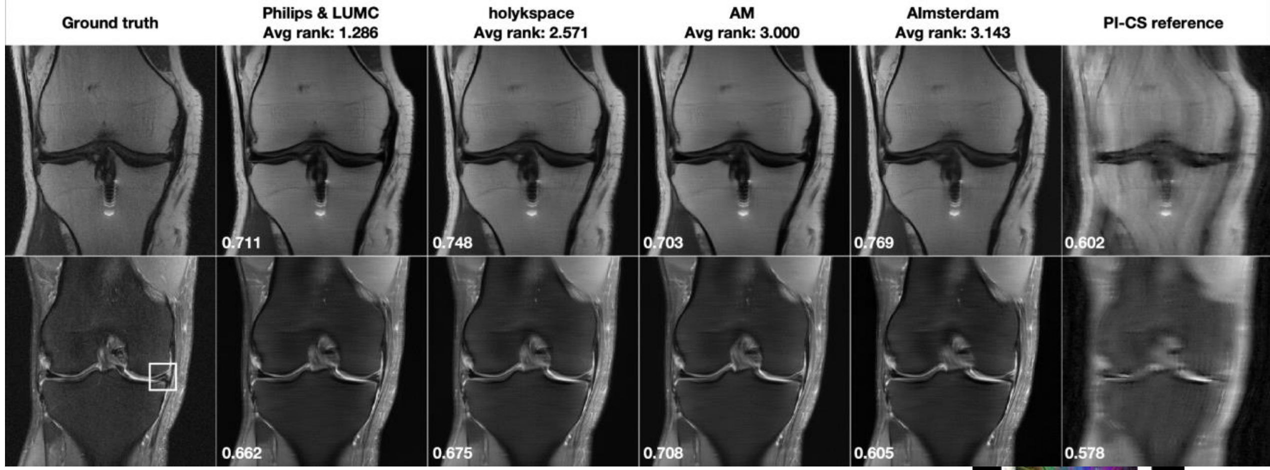
Undersampled MRI

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)

HELMHOLTZ IMAGING

Learning in Image Reconstruction

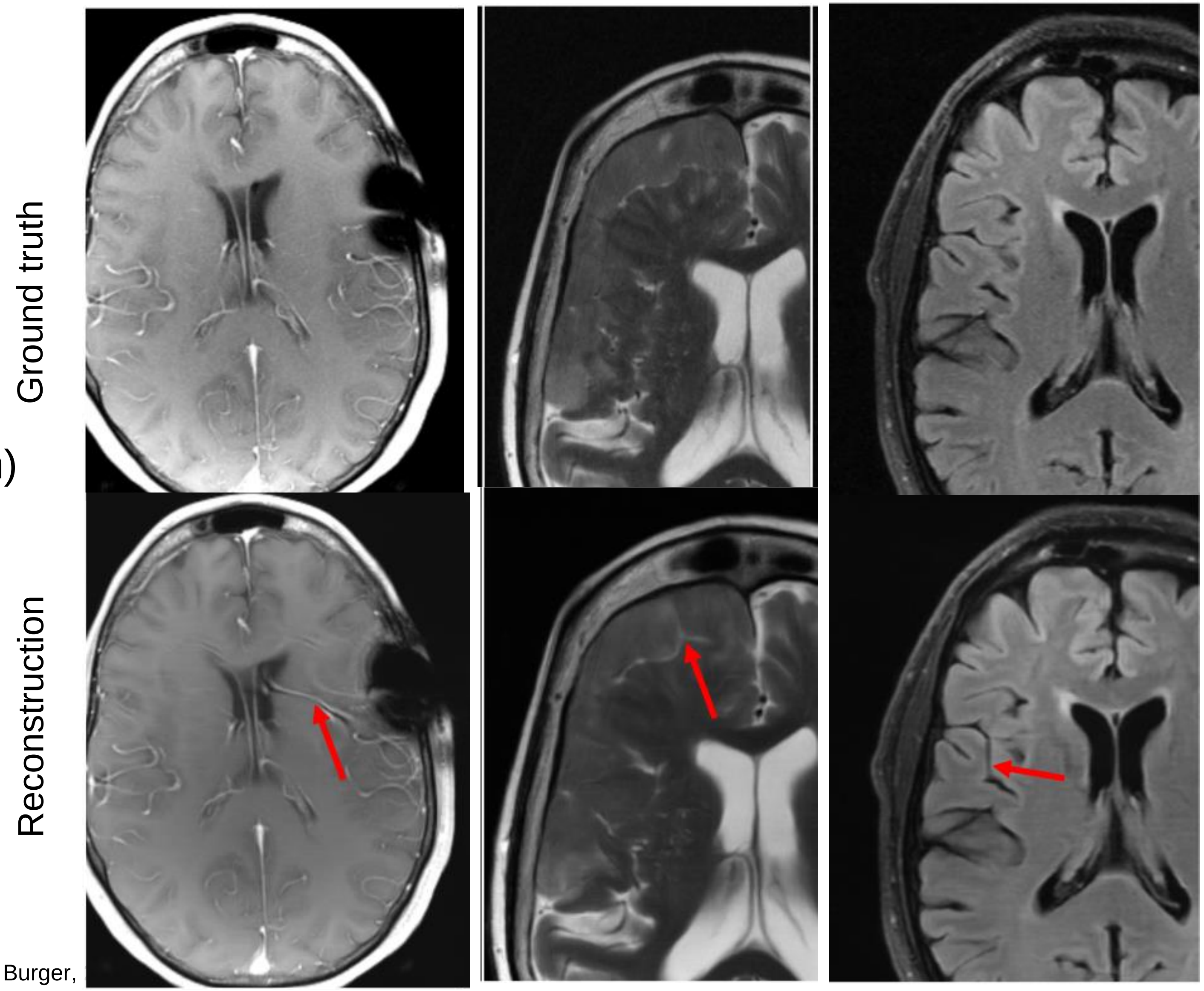
Undersampled MRI

Majority of results convincing

But possible hallucinations
on few data sets

Not recognizable by
experienced radiologists

(courtesy Florian Knoll, Erlangen)



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

HELMHOLTZ IMAGING

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

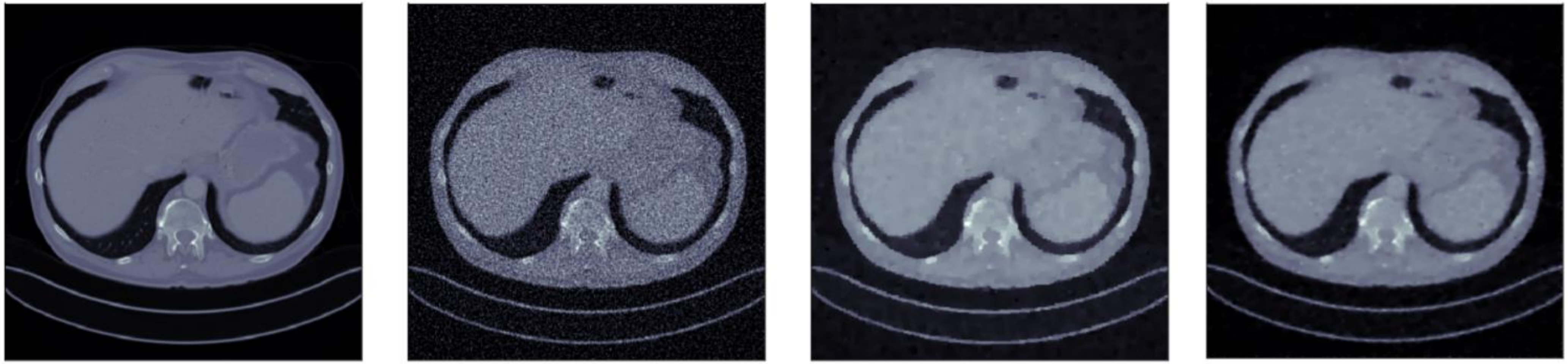
Learned Regularizers

Adversarial regularization with source condition

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

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

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

HELMHOLTZ IMAGING

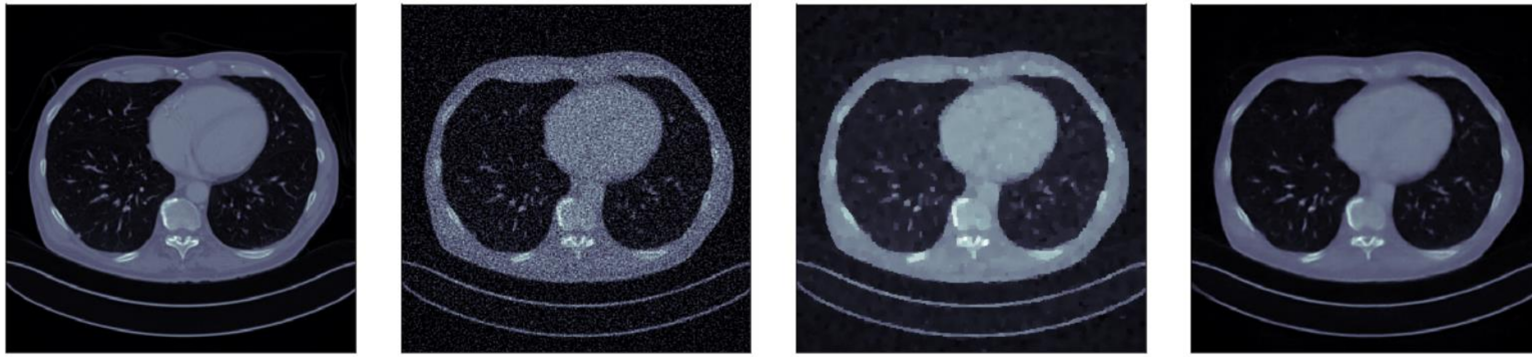
Learned Regularizers

Adversarial regularization with source condition

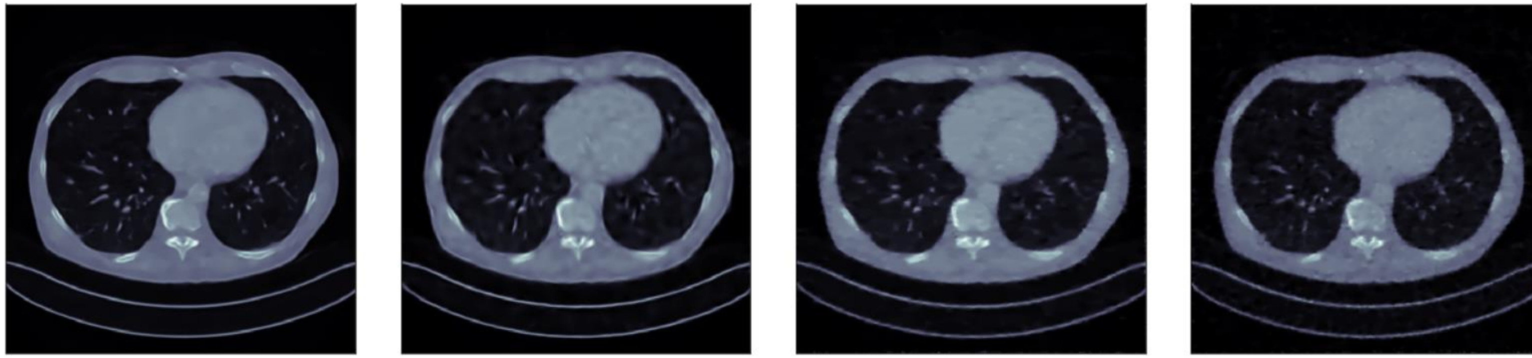
method	PSNR (dB)	SSIM	# param.	reconstruction time (ms)
FBP	21.28 ± 0.13	0.20 ± 0.02	1	37.0 ± 4.6
TV	30.31 ± 0.52	0.78 ± 0.01	1	28371.4 ± 1281.5
Supervised methods				
U-Net	34.50 ± 0.65	0.90 ± 0.01	7215233	44.4 ± 12.5
LPD	35.69 ± 0.60	0.91 ± 0.01	1138720	279.8 ± 12.8
Unsupervised methods				
AR	33.84 ± 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

Best case comparison:
Supervised learning and methods with less constraints superior

However, more interpretability and robustness with source condition constraint



(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, 0.8

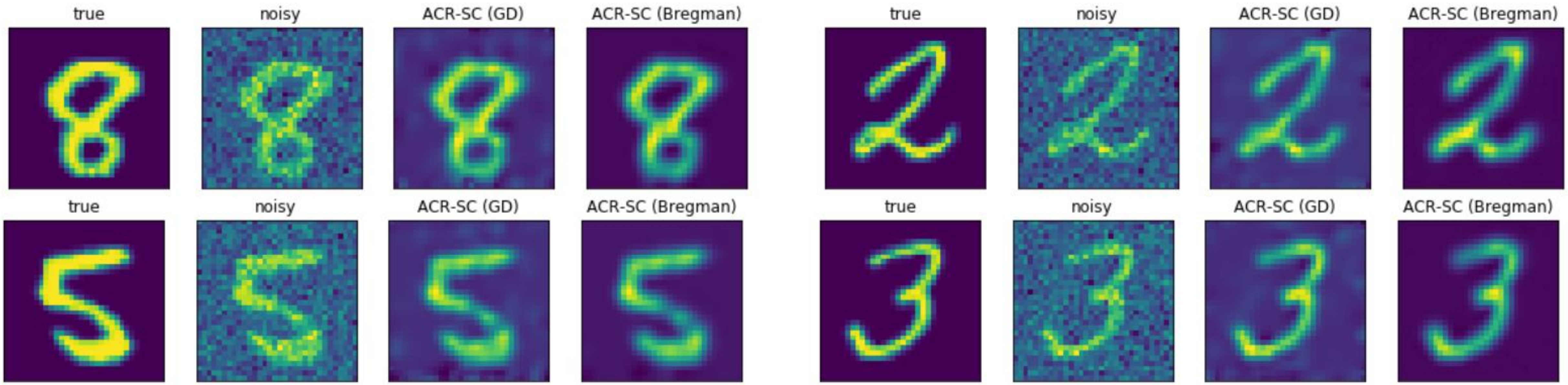
Learned Regularizers

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]

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



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 surpringly 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†

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 Öktem[4], Marcelo Pereyra[5], and Carola-Bibiane Schönlieb[1]

Learning the optimal Tikhonov regularizer for inverse problems

Giovanni S. Alberti[1], Ernesto De Vito[1], Matti Lassas[2], Luca Ratti[1], Matteo Santacesaria[1]

OPEN ACCESS
IOP Publishing — Inverse Problems
Inverse Problems 38 (2022) 115005 (21pp) https://doi.org/10.1088/1361-6420/ac9011

Regularization theory of the analytic deep prior approach

Clemens Arndt*

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*

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

Setup: noise independent of u, unbiased

$$f^\delta = Ku + \eta \qquad \mathbb{E}(\eta) = 0$$

Learning in Image Reconstruction

A way to theory

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

$$R(f^\delta;\ g^\delta) = \sum_n g^\delta(\sigma_n)\langle f^\delta, v_n\rangle u_n$$

$$\overline{g} = \arg\min_g \ \mathbb{E}_{u,\nu}(\|u - R(f^\delta;\ g)\|^2) \qquad\qquad g_n = g(\sigma_n)$$

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)

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

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

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

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

$$= \sum_n (1-\sigma_n g_n)^2\, \Pi_n + g_n^2 \Delta_n,$$

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

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

Leads again to effective regularization

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

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

HELMHOLTZ IMAGING

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

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

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

$$\Delta_n \sim \delta \qquad \Pi_n \to 0$$

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

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

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

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

HELMHOLTZ IMAGING

Learning in Image Reconstruction

Expected smoothness of reconstructions

Define

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

Can be computed efficiently

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

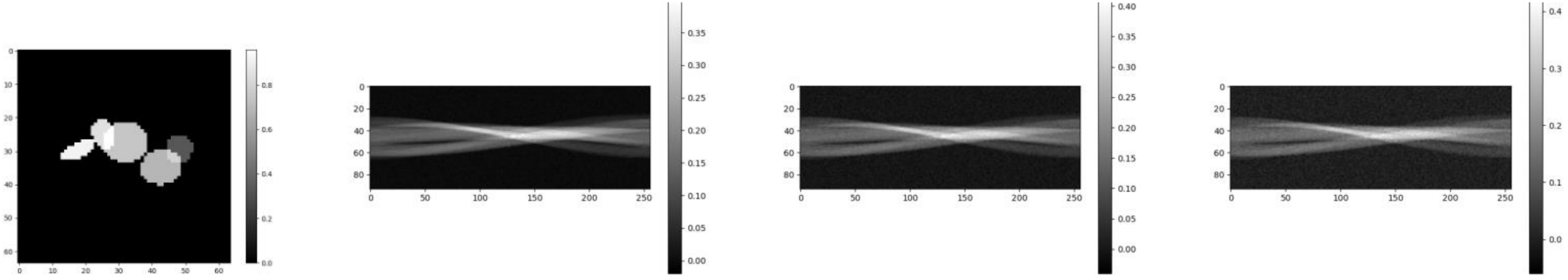
Note: reconstructed solutions are always smoother than clean training data (oversmoothed)

Explanation from regularization applied to rough noise

Learning in Image Reconstruction

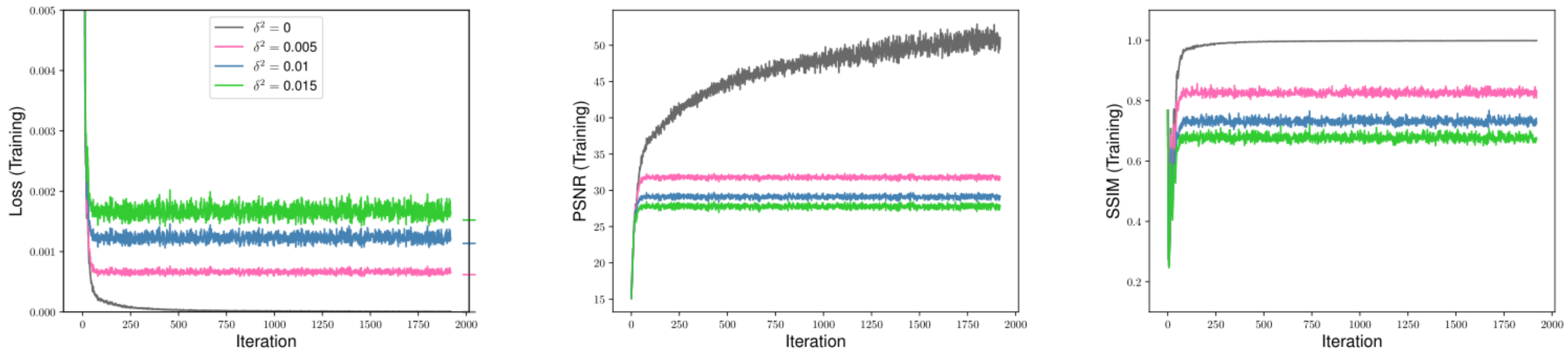
Spectral regularization in tomography

Training data for different noise level



(a) $\mathbf{u}$ (b) $\delta^2 = 0.005$ (c) $\delta^2 = 0.01$ (d) $\delta^2 = 0.015$

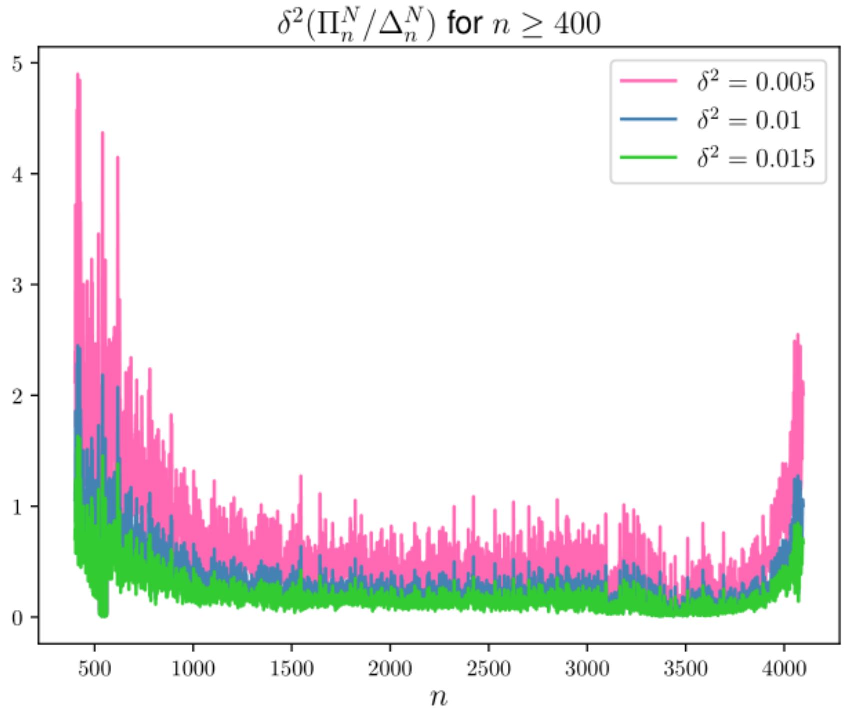
Training statistics

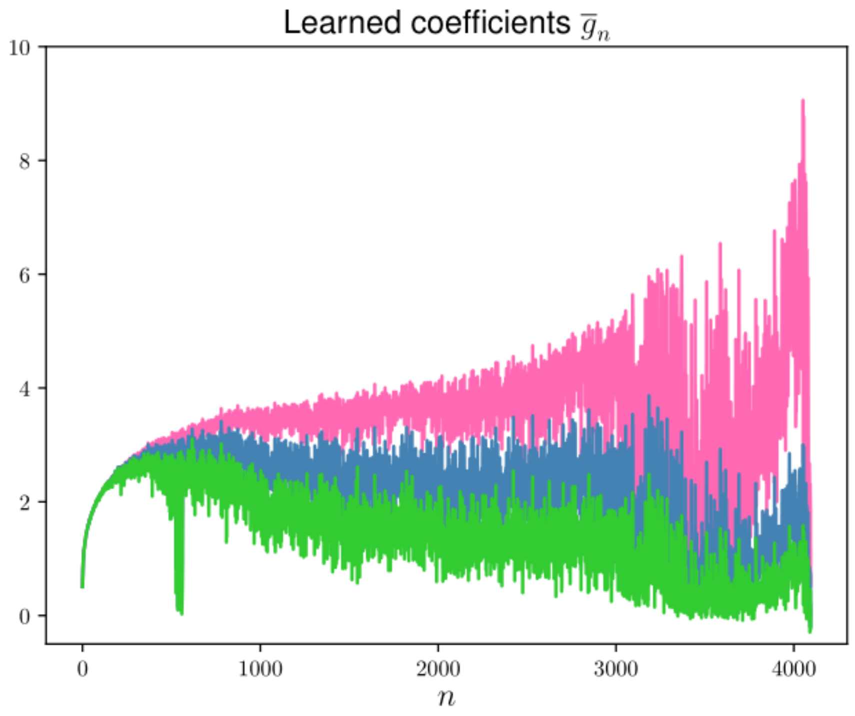


Learning in Image Reconstruction

Spectral regularization in tomography

Oversmoothing in computational results



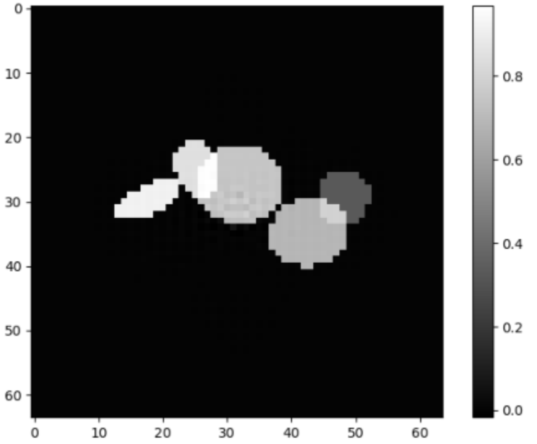


HELMHOLTZ IMAGING

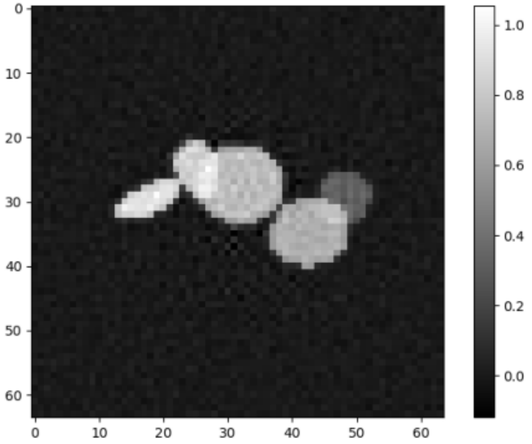
Learning in Image Reconstruction

Spectral regularization in tomography

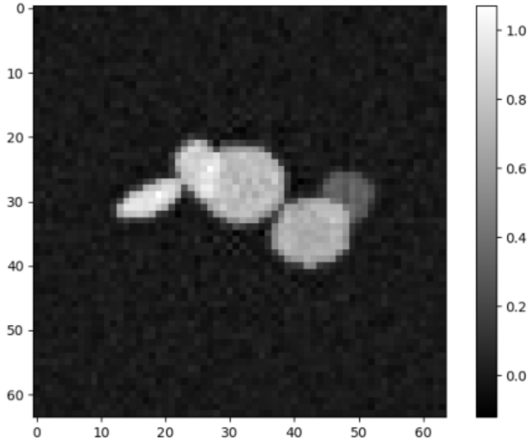
Typical reconstructions



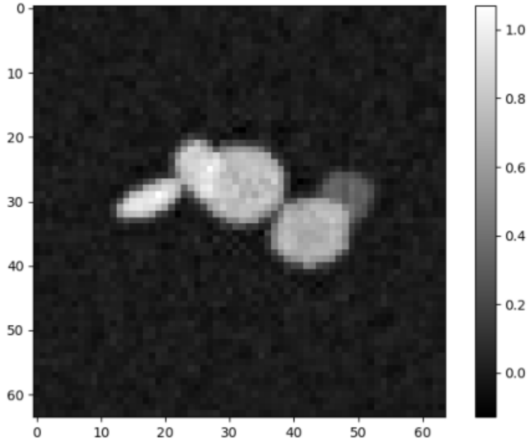
(a) $\delta^2 = 0$



(b) $\delta^2 = 0.005$



(c) $\delta^2 = 0.01$



(d) $\delta^2 = 0.015$

HELMHOLTZ IMAGING

Credits

Samira Kabri, DESY

Martin Benning, Danilo Riccio, QMU

Carola Schönlieb, Subhadip Mukherjee, Cambridge

Michael Möller, Alexander Auras, Siegen

HELMHOLTZ IMAGING

www.helmholtz-imaging.de

European Research Council
Established by the European Commission

Deutsche Forschungsgemeinschaft DFG

NoMADS

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Convergence of learned regularization

Consider a sequence of learned regularizations with respect to noise level

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

Assumptions:

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

$$\delta^2 = \max_n \mathit{\Delta}_n.$$

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

$$\mathit{\Delta}_n \geq c\, \delta^2\, \mathit{\Pi}_n$$

for all $n \geq n_0$.

(3) The sequence of $\mathit{\Pi}_n$ is summable, i.e.,

$$\sum_n \mathit{\Pi}_n < \infty.$$

Convergence of learned regularization

Consider a sequence of learned regularizations with respect to noise level

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

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

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

as $\delta \to 0$.

HELMHOLTZ IMAGING

Supervised learning

Error

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

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

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