UNIVERSITY OF CAMBRIDGE

Variational regularisation for inverse problems with imperfect forward operators and general noise models

Leon Bungert[1], Martin Burger[1], Yury Korolev[2] and Carola Schönlieb[2]

[1]Department Mathematik, University of Erlangen-Nürnberg, Germany
[2]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK

SIAM Conference on Imaging Science 15 July 2020

Layout

Introduction

Convergence Analysis

Discrepancy Principle

Inverse Problems

Inverse problem:

$$Au = \bar{f},$$

- $A\colon U \to F$ is the forward operator (linear in this talk),
- $\bar{f} \in U$ exact (unattainable) data,
- f^δ noisy measurement with amount of noise characterised by $\delta > 0$.

Variational Regularisation

Variational regularisation:

$$\min_{u \in U} \frac{1}{\alpha} \mathcal{H}(Au \mid f^\delta) + \mathcal{J}(u),$$

- $\mathcal{H}(\cdot \mid f^\delta)$ is the *fidelity function* that models the noise (e.g., Kullback-Leibler divergence, L^p-norm, Wasserstein distance),
- $\mathcal{J}(\cdot)$ is the regularisation term (e.g., Total Variation, ℓ^1-norm),
- α is the regularisation parameter.

Imperfect Forward Operators

Forward operator $A\colon U \to F$ often
- is not perfectly known (errors in geometry, coefficients of a PDE, convolution kernel), or
- can only be evaluated approximately (simplified models, discretisation errors).

Regularisation under operator errors:

- Goncharskii, Leonov, Yagola (1973). A generalized discrepancy principle;
- Hofmann (1986). Optimization aspects of the generalized discrepancy principle in regularization;
- Neubauer, Scherzer (1990). Finite-dimensional approximation of Tikhonov regularized solutions of nonlin. ill-posed prob.;
- Pöschl, Resmerita, Scherzer (2010). Discretization of variational regularization in Banach spaces;
- Bleyer, Ramlau (2013). A double regularization approach for inverse problems with noisy data and inexact operator;
- YK, Yagola (2013). Making use of a partial order in solving inverse problems;
- YK (2014). Making use of a partial order in solving inverse problems: II;
- YK, Lellmann (2018). Image reconstruction with imperfect forward models and applications in deblurring;
- Burger, YK, Rasch (2019). Convergence rates and structure of solutions of inv. prob. with imperfect forward models;
- Dong et al. (2019). Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector;

Bayesian approximation error modelling:

- Kaipio, Somersalo (2005). Statistical and computational inverse problems;
- Arridge et al. (2006). Approximation errors and model reduction with an application in optical diffusion tomography;
- Hansen et al. (2014). Accounting for imperfect forw. model. in geophys. inv. prob. - exemplified for crosshole tomography;
- Calvetti et al. (2018). Iterative updating of model error for Bayesian inversion;
- Rimpiläinen et al. (2019). Improved EEG source localization with Bayes. uncert. modelling of unknown skull conductivity;
- Riis, Dong, Hansen (2020). Computed tomography reconstr. with uncert. view angles by iter. updated model discrepancy.

Learned Forward Operators

Forward operator (or a correction to it) is learned from training pairs

$$(u^i, f^i)_{i=1}^n \quad \text{s.t.} \quad Au^i = f^i.$$

Learned forward operators:

- Aspri, YK, Scherzer (2019). Data-driven regularisation by projection;
- Bubba et al. (2019). Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography;
- Schwab, Antholzer, Haltmeier (2019). Deep null space learning for inverse problems: convergence analysis and rates;
- Boink, Brune (2019). Learned SVD: solving inverse problems via hybrid autoencoding;
- Lunz et al. (2020). On learned operator correction;
- Nelsen, Stuart (2020). The random feature model for input-output maps between Banach spaces.

Contribution: combining general fidelities and operator errors

Variational regularisation with exact operator

$$\min_{u \in U} \frac{1}{\alpha} \mathcal{H}(Au \mid f^\delta) + \mathcal{J}(u).$$

Modelling operator error using partial order in a Banach lattice

$$A^l \leqslant A \leqslant A^u \quad \text{(in a sense made precise later).}$$

Proposed: variational regularisation with interval operator

$$\min_{\substack{u \in U \\ v \in F}} \frac{1}{\alpha} \mathcal{H}(v \mid f^\delta) + \mathcal{J}(u) \quad \text{s.t.} \quad A^l u \leqslant_F v \leqslant_F A^u u.$$

- Convergence rates for a priori choices of α (depending on δ and $\|A^u - A^l\|$);
- Convergence rates for a posteriori choices of α (discrepancy principle; depending on δ, f^δ, A^l and A^u).

Bungert, Burger, YK, Schönlieb (2020). Variational regularisation for inverse problems with imperfect forward operators and general noise models. arXiv:2005.14131

Layout

Banach Lattices

- Vector space $\mathcal{X}$ with partial order $\leqslant$ called an *ordered vector space* if

$$x \leqslant y \implies x + z \leqslant y + z \qquad \forall\ x, y, z \in \mathcal{X},$$
$$x \leqslant y \implies \lambda x \leqslant \lambda y \qquad \forall\ x, y \in \mathcal{X} \text{ and } \lambda \in \mathbb{R}_+.$$

Banach Lattices

- Vector space $\mathcal{X}$ with partial order $\leqslant$ called an *ordered vector space* if

$$x \leqslant y \implies x + z \leqslant y + z \qquad \forall\ x, y, z \in \mathcal{X},$$
$$x \leqslant y \implies \lambda x \leqslant \lambda y \qquad \forall\ x, y \in \mathcal{X} \text{ and } \lambda \in \mathbb{R}_+.$$

- A *vector lattice* (or a *Riesz space*) is an ordered vector space $\mathcal{X}$ with well defined suprema and infima

$$\forall x, y \in \mathcal{X} \quad \exists\, x \vee y \in \mathcal{X},\ x \wedge y \in \mathcal{X};$$
$$x \vee 0 = x_+, \quad (-x)_+ = x_-, \quad x = x_+ - x_-, \quad |x| = x_+ + x_-.$$

Banach Lattices

- Vector space $\mathcal{X}$ with partial order $\leqslant$ called an *ordered vector space* if

$$x \leqslant y \implies x + z \leqslant y + z \qquad \forall\ x, y, z \in \mathcal{X},$$
$$x \leqslant y \implies \lambda x \leqslant \lambda y \qquad \forall\ x, y \in \mathcal{X} \text{ and } \lambda \in \mathbb{R}_+.$$

- A *vector lattice* (or a *Riesz space*) is an ordered vector space $\mathcal{X}$ with well defined suprema and infima

$$\forall x, y \in \mathcal{X} \quad \exists\, x \vee y \in \mathcal{X},\ x \wedge y \in \mathcal{X};$$
$$x \vee 0 = x_+, \quad (-x)_+ = x_-, \quad x = x_+ - x_-, \quad |x| = x_+ + x_-.$$

- A *Banach lattice* is a vector lattice $\mathcal{X}$ with a monotone norm, i.e.

$$\forall x, y \in \mathcal{X} \quad |x| \geqslant |y| \implies \|x\| \geqslant \|y\|.$$

Banach Lattices

- Vector space $\mathcal{X}$ with partial order $\leqslant$ called an *ordered vector space* if

$$x \leqslant y \implies x + z \leqslant y + z \qquad \forall\ x, y, z \in \mathcal{X},$$
$$x \leqslant y \implies \lambda x \leqslant \lambda y \qquad \forall\ x, y \in \mathcal{X} \text{ and } \lambda \in \mathbb{R}_+.$$

- A *vector lattice* (or a *Riesz space*) is an ordered vector space $\mathcal{X}$ with well defined suprema and infima

$$\forall x, y \in \mathcal{X} \quad \exists\, x \vee y \in \mathcal{X},\ x \wedge y \in \mathcal{X};$$
$$x \vee 0 = x_+, \quad (-x)_+ = x_-, \quad x = x_+ - x_-, \quad |x| = x_+ + x_-.$$

- A *Banach lattice* is a vector lattice $\mathcal{X}$ with a monotone norm, i.e.

$$\forall x, y \in \mathcal{X} \quad |x| \geqslant |y| \implies \|x\| \geqslant \|y\|.$$

- Partial order for linear operators $A, B \colon \mathcal{X} \to \mathcal{Y}$ is defined as

$$A \geqslant B \quad \text{if} \quad \forall x \geqslant 0 \text{ in } \mathcal{X} \quad \implies \quad Ax \geqslant Bx \text{ in } \mathcal{Y}.$$

Convergence of the Data and the Operator

We consider sequences

$$\begin{aligned}
A_n^l, A_n^u\colon \quad & A_n^l \leqslant A \leqslant A_n^u \quad \forall n, \\
& \|A_n^u - A_n^l\| \leqslant \eta_n \to 0 \quad \text{as } n \to \infty, \\
f_n, \delta_n\colon \quad & \mathcal{H}(\bar{f} \mid f_n) \leqslant \delta_n \quad \forall n, \\
& \delta_n \to 0 \quad \text{as } n \to \infty, \\
\alpha_n\colon \quad & \alpha_n \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Sequence of corresponding primal solutions

$$(u_n, v_n), \quad n = 1, ..., \infty.$$

General Estimate

Assumption (Source condition)

There exists $\omega^\dagger \in F^*$ *s.t.*

$$A^*\omega^\dagger \in \partial\mathcal{J}(u^\dagger_{\mathcal{J}}).$$

Theorem (Bungert, Burger, YK, Schönlieb'20)

Under standard assumptions the following estimate holds for the Bregann distance $D_{\mathcal{J}}(u_n, u^\dagger_{\mathcal{J}})$ *between the approximate solution* u_n *and the* $\mathcal{J}$*-minimising solution* $u^\dagger_{\mathcal{J}}$

$$D_{\mathcal{J}}(u_n, u^\dagger_{\mathcal{J}}) \leqslant \frac{\delta_n}{\alpha_n} + \frac{1}{\alpha_n}[\mathcal{H}^*(\alpha_n\omega^\dagger \mid f_n) - \langle \alpha_n\omega^\dagger, \bar{f}\rangle] + C\eta_n.$$

φ-divergences

Definition

Let $\varphi\colon (0,\infty) \to \mathbb{R}_+$ be convex and $\varphi(1) = 0$. For $\rho, \nu \in \mathcal{P}(\Omega)$ with $\rho \ll \nu$ the φ-divergence is defined as follows

$$d_\varphi(\rho \mid \nu) := \int_\Omega \varphi\left(\frac{\mathrm{d}\rho}{\mathrm{d}\nu}\right) \mathrm{d}\nu.$$

We further assume that $\varphi^*(x) = x + r(x)$, where φ^* is the convex conjugate and $r(x)/x \to 0$ as $x \to 0$.

- Kullback-Leibler divergence: $\varphi(x) = x\log(x) + x - 1$;
- χ^2 divergence: $\varphi(x) = (x-1)^2$;
- Squared Hellinger distance: $\varphi(x) = (\sqrt{x}-1)^2$;
- Total variation: $\varphi(x) = |x-1|$.

φ-divergences

Theorem (Bungert, Burger, YK, Schönlieb'20)

Under standard assumptions the following convergence rate holds

$$D_{\mathcal{J}}(u_n, u_{\mathcal{J}}^\dagger) = O\left(\frac{\delta_n}{\alpha_n} + \frac{r(\alpha_n)}{\alpha_n} + \eta_n\right).$$

For an optimal choice of α we get
- Kullback-Leibler divergence, χ^2 divergence, Squared Hellinger distance:

$$D_{\mathcal{J}}(u_n, u_{\mathcal{J}}^\dagger) = O\left((\delta_n)^{\frac{1}{2}} + \eta_n\right);$$

- Total variation:

$$D_{\mathcal{J}}(u_n, u_{\mathcal{J}}^\dagger) = O\left(\delta_n + \eta_n\right) \quad \text{(exact penalisation).}$$

Strongly Coercive Fidelities

Theorem (Bungert, Burger, YK, Schönlieb'20)

Suppose that the fidelity function $\mathcal{H}$ satisfies

$$\frac{1}{\lambda}\|v - f\|_F^\lambda \leqslant \mathcal{H}(v \mid f)$$

for all $v, f \in F$, where $\lambda \geqslant 1$. Then under standard assumptions and for an optimal choice of α the following rate holds

$$D_{\mathcal{J}}(u_n, u_{\mathcal{J}}^\dagger) \quad = \quad O\left(\delta_n^{\frac{1}{\lambda}} + \eta_n\right).$$

- Powers of norms;
- Wasserstein distances (coercive in the Kantorovich-Rubinstein norm).

Mixed Noise

Sum of fidelities:

$$\mathcal{H}(v \mid f) = \mathcal{H}_1(v \mid f) + \mathcal{H}_2(v \mid f),$$
$$\implies D_{\mathcal{J}}(u_n, u^\dagger_{\mathcal{J}}) = O\left(\left(R_1(\cdot, \delta_n) \,\square\, R_2(\cdot, \delta_n) \right) (\alpha_n) + \eta_n \right),$$

where $R_{1,2}(\cdot, \delta_n)$ are individual rates.

- Hintermüller, Langer (2013). Subspace correction methods for a class of nonsmooth and nonadditive convex variational problems with mixed L^1/L^2 data-fidelity in image processing;
- Yue et al. (2014). A locally adaptive L1-L2 norm for multiframe super-resolution of images with mixed noise and outliers;
- Langer (2017). Automated parameter selection in the-TV model for removing Gaussian plus impulse noise.

Infimal convolution of fidelities:

$$\mathcal{H}(v \mid f) = (\mathcal{H}_1(\cdot \mid 0) \,\square\, \mathcal{H}_2(\cdot \mid f))(v),$$
$$\implies D_{\mathcal{J}}(u_n, u^\dagger_{\mathcal{J}}) = O\left(R_1(\alpha_n, \delta_n) + R_2(\alpha_n, \delta_n) + \eta_n \right).$$

- Calatroni, De Los Reyes, Schönlieb (2017). Infimal convolution of data discrepancies for mixed noise removal.

Layout

Discrepancy Principle

Discrepancy principle for for exact operators

$$\alpha_n = \sup\{\alpha > 0 \colon \|Au_n^{\alpha_n} - f_n\|^2 \leqslant \tau\delta_n\}.$$

- Morozov (1966). On the solution of functional equations by the method of regularisation;
- Bonesky (2008). Morozov's discrepancy principle and Tikhonov-type functionals;
- Anzengruber, Ramlau (2009). Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators;
- Sixou, Hohweiller, Ducros (2018). Morozov principle for Kullback-Leibler residual term and Poisson noise.

Generalisation to errors in the operator (in the Hilbert space setting)

$$\alpha_n = \sup\{\alpha > 0 \colon \|Au_n^{\alpha_n} - f_n\|^2 = (\sqrt{\delta_n} + h_n\|u_n^{\alpha_n}\|)^2\}.$$

- Goncharskii, Leonov, Yagola (1973). A generalized discrepancy principle;
- Hofmann (1986). Optimization aspects of the generalized discrepancy principle in regularization;
- Lu et al. (2010). On the generalized discrepancy principle for Tikhonov regularization in Hilbert scales.

We propose

$$\alpha_n = \sup\{\alpha > 0 \colon \mathcal{H}(v_n^\alpha \mid f_n) \leqslant \tau\delta_n\},$$

where (u_n^α, v_n^α) solve

$$\min_{u,v} \frac{1}{\alpha}\mathcal{H}(v \mid f^{\delta_n}) + \mathcal{J}(u) \quad \text{s.t.} \quad A_n^l u \leqslant v \leqslant A_n^u u.$$

Discrepancy Principle

Theorem (Bungert, Burger, YK, Schönlieb'20)

Under standard assumptions, for strongly coercive fidelities

$$D_{\mathcal{J}}(u_n^{\alpha_n}, u_{\mathcal{J}}^{\dagger}) = O\left(\delta_n^{\frac{1}{\lambda}} + \eta_n\right),$$

i.e. we recover optimal rates. If the φ-divergence satisfies Pinsker's inequality, we also recover optimal rates.
E.g., for the Kullback-Leibler divergence Pinsker's inequality says

$$\|\bar{f} - f_n\| \leqslant \sqrt{2\mathcal{H}(\bar{f} \mid f_n)} = O(\sqrt{\delta_n}),$$

hence

$$D_{\mathcal{J}}(u_n^{\alpha_n}, u_{\mathcal{J}}^{\dagger}) = O(\sqrt{\delta_n} + \eta_n).$$

Conclusions

- Convergence rates for variational regularization in Banach lattices for problems with imperfect forward operators and general fidelity functions:
  - norm-type fidelities;
  - Wasserstein distances;
  - φ-divergences, e.g. Kullback-Leibler;
  - mixed noise;
- recover optimal rates for problems with exact operator;
- extend the discrepancy principle to a combination of an inexact operator and a general fidelity;
  - also recover optimal rates;
- a general and versatile approach to problems with complicated measurement noise and inexact modelling.

Bungert, Burger, YK, Schönlieb (2020). Variational regularisation for inverse problems with imperfect forward operators and general noise models. arXiv:2005.14131

So long, and thanks for all the funding

Unterstützt von / Supported by
Alexander von Humboldt
Stiftung/Foundation

FUNDED BY
THE ROYAL SOCIETY

CCIMI
CANTAB CAPITAL INSTITUTE FOR THE MATHEMATICS OF INFORMATION

NPL
National Physical Laboratory

HUGHES HALL
UNIVERSITY OF CAMBRIDGE