22. Oct. 2023

universität innsbruck
Institut für Mathematik

Preprint-Series: Department of Mathematics - Applied Mathematics

Data-driven Morozov regularization of inverse problems

M.Haltmeier, R. Kowar, M. Tiefentaler

AppliedMathematics

Technikerstraße 13 - 6020 Innsbruck - Austria
Tel.: +43 512 507 53803 Fax: +43 512 507 53898
https://applied-math.uibk.ac.at

Data-driven Morozov regularization of inverse problems

Markus Haltmeier, Richard Kowar, and Markus Tiefentaler

Department of Mathematics, University of Innsbruck
Technikerstrasse 13, 6020 Innsbruck, Austria

October 22, 2023

Abstract

The solution of inverse problems is central to a wide range of applications including medicine, biology, and engineering. These problems require finding a desired solution in the presence of noisy observations. A key feature of inverse problems is their ill-posedness, which leads to unstable behavior under noise when standard solution methods are used. For this reason, regularization methods have been developed that compromise between data fitting and prior structure. Recently, data-driven variational regularization methods have been introduced, where the prior in the form of a regularizer is derived from provided ground truth data. However, these methods have mainly been analyzed for Tikhonov regularization, referred to as Network Tikhonov Regularization (NETT). In this paper, we propose and analyze Morozov regularization in combination with a learned regularizer. The regularizers, which can be adapted to the training data, are defined by neural networks and are therefore non-convex. We give a convergence analysis in the non-convex setting allowing noise-dependent regularizers, and propose a possible training strategy. We present numerical results for attenuation correction in the context of photoacoustic tomography.

Key words: Inverse problems, learned regularizer, convergence analysis, Morozov regularization.

MSC codes: 65F22; 68T07

1 Introduction

In this paper, we consider the solution of linear inverse problems where we aim to reconstruct the unknown $x \in \mathbb{X}$ from noisy data

$$y_\delta = \mathbf{A}x + \eta_\delta \,. \tag{1.1}$$

Here $\mathbf{A}\colon \mathbb{X} \to \mathbb{Y}$ is a linear bounded operator between Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$, and η_δ is the data error that satisfies $\|\eta_\delta\| \leq \delta$ with noise level $\delta \geq 0$. We are especially interested in the ill-posed case where solving (1.1) without prior information is non-unique or unstable. Several applications in medical image reconstruction, nondestructive testing, and remote sensing are instances of such linear inverse problems [9, 23, 28].

Characteristic features of inverse problems are the non-uniqueness of solutions and the unstable dependence of solutions on data perturbations. To account for these two issues, one must apply regularization methods (see, for example, [4, 9, 14–16, 21, 28, 29, 31]) that serve two main purposes: First, in the case of exact data $y \in \operatorname{ran}(\mathbf{A})$, they select a specific solution $\mathbf{B}_0(y)$ among all possible solutions of the exact data equation $y = \mathbf{A}x$. Second, to account for noise, they define stable approximations to $\mathbf{B}_0$ in the form of continuous mappings $\mathbf{B}_\alpha\colon \mathbb{Y} \to \mathbb{X}$ that converge to $\mathbf{B}_0$ as $\alpha \to 0$ in an appropriate sense.

1.1 Morozov regularization

There are several well established methods for the stable solution of inverse problems. A general class of regularization methods are variational regularization methods which includes Tikhhonov regularization, Ivanov regularization (the method of quasi solutions), and Morozov regularization (the residual method) as special cases. In Tikhonov regularization, approximate solutions are defined as minimizers of $\|\mathbf{A}x - y_\delta\|^2/2 + \alpha\mathcal{R}(x)$, where $\mathcal{R}\colon \mathbb{X} \to [0, \infty]$ is a regularization functional that measures the feasibility of a potential solution and α is the regularization parameter. Ivanov regularization considers minimizers of $\|\mathbf{A}x - y_\delta\|$ over the set $\{x \in \mathbb{X} \mid \mathcal{R}(x) \leq \tau\}$ for some $\tau > 0$. In this paper, we consider Morozov regularization where approximate solutions defined as solutions of

$$\min_{x \in \mathbb{X}} \mathcal{R}(x) \quad \text{s.t.} \quad \|\mathbf{A}x - y_\delta\| \leq \delta\,. \tag{1.2}$$

Compared to Tikhonov regularization and Ivanov regularization, the latter has the advantage that no additional regularization parameter has to be selected, which is typically a difficult issue. Relations between Tikhonov regularization, Ivanov regularization and Morozov regularization are carefully studied in [16].

Note that variational regularization methods are designed to approximate $\mathcal{R}$-minimizing solution of $\mathbf{A}x = y$ for the limit $\delta \to 0$, defined as elements in $\arg\min\{\mathcal{R}(x) \mid \mathbf{A}x = y\}$. This addresses the non-uniqueness in the case of exact data. To account for noise, Morozov regularization relaxes the strict data consistency $\mathbf{A}x = y$ to data proximity $\|\mathbf{A}x - y_\delta\| \leq \delta$. Even in the case that $\mathbf{A}$ is injective, different regularization terms behave differently and significantly affect convergence. Therefore the choice of the regularizer is crucial and a nontrivial issue. Classical choices for the regularizers are the squared Hilbert space norm $\mathcal{R}(x) = \|x\|^2/2$ or

the ℓ^1-norm $\mathcal{R}(x) = \sum_{\lambda\in\Lambda}|\langle u_\lambda, x\rangle|$, where $(u_\lambda)_\lambda$ is a frame of $\mathbb{X}$. These regularizers may not be optimally adapted to highly structured signal classes, as is often the case in practical applications. In this paper, we address this issue and propose a data-driven regularizer using neural networks adapted to the signal class represented by training data in combination with Morozov regularization.

1.2 Neural network regularizers

In this paper, we study Morozov regularization with a Neural network based and data-driven and noise-dependent regularizer

$$\mathcal{R}_\delta(x) = \frac{1}{2}\|\mathbf{\Phi}_{\theta(\delta)}(x) - x\|^2 + \lambda\,\mathcal{Q}(x)\,. \tag{1.3}$$

Here $\mathbf{\Phi}_{\theta(\delta)}\colon \mathbb{X} \to \mathbb{X}$ is a neural network tuned to noisy data, and $\mathcal{Q}\colon \mathbb{X} \to [0,\infty]$ is an additional noise-adaptive data-driven regularization term. We will refer to (1.2) with the data-driven regularizer (1.3) instead of the fixed regularizer $\mathcal{R}$ as data-driven Morozov (DD-Morozov) regularization and show that, under reasonable assumptions, this gives a regularization method. Furthermore, we present a training strategy for selecting the noise-dependent neural network among a given architecture.

Besides stabilizing the signal reconstruction, the main purpose of a particular regularizer is to fit the reconstructions to a certain set where the true signals are likely to be contained. In reality, this set is not known analytically, but it is possible to draw examples from it. For this reason, we follow the learning paradigm and choose training signals $x_1, \dots, x_N \in \mathbb{X}$ and adapt the architecture $(\mathbf{\Phi}_\theta)_{\theta\in\theta}$ to x_i. More precisely, $\theta = \theta(\delta)$ is chosen such that $\mathbf{\Phi}_\theta(x_i) \simeq x_i$ and $\mathbf{\Phi}_\theta(x_i + r_{i,j}) \simeq x_i$, where $r_{i,j}$ are appropriate perturbations. Thus, $\mathcal{R}_\delta$ has small values for the exact x_i and larger values for the perturbed signals $x_i + r_{i,j}$. Since the training data is only taken from a certain subset of $\mathbb{X}$, it is difficult to obtain the necessary coercive condition from training alone. Therefore, it seems natural to add another regularization term $\mathcal{Q}$ which is known to be coercive.

While learned regularizers have recently become popular in the context of Tikhonov regularization [11, 19, 20, 24], we are not aware of any work utilizing the Morozov variant. In fact, our analysis as well as the training strategy are closely related to the Network Tikhonov Approach (NETT) of [19, 24]. However, in this paper we use a refined training strategy that is suitable for ill-posed problems and we consider a different regularization concept. A different strategy for learning a network regularizer has been proposed in [20] in the context of adversarial regularization. Other approaches for learning a regularizer are the fields of experts model [27], deep total variation [17] or ridge regularizers [11]. Other data-driven regularization methods for inverse problems can be found for example in [1–3, 7, 14, 22, 25, 30] and the

references therein. From the theoretical side, Morozov regularization in a general non-convex context has been studied in [12]. The analysis we present below allows the regularizer to be noise-dependent and further we derive strong convergence under total nonlinearity condition of [19].

1.3 Outline

The remainder of this paper is organized as follows. In Section 2 we present our theoretical results. In particular we present the convergence analysis (Section 2.1) and the prosed training strategy (Section 2.2). In Section 3 we present numerical results illustrating our proposal. Specifically, we test our reconstruction strategy on a severely ill-posed problem regarding attenuation correction for photoacoustic tomography in damping media [18]. To numerically solve (1.2) we implement the primal dual scheme of [6]. The paper concludes with a short summary in Section 4 .

2 Theory

Throughout this paper, $\mathbb{X}$, $\mathbb{Y}$ are Hilbert spaces and $\mathbf{A}\colon \mathbb{X} \to \mathbb{Y}$ a bounded linear operator. Recall that a functional $\mathcal{R}\colon \mathbb{X} \to [0,\infty]$ is coercive, if $\mathcal{R}(x_n) \to \infty$ for all sequences $(x_n)_{n\in\mathbb{N}} \in \mathbb{X}^{\mathbb{N}}$ with $\|x_n\|_{\mathbb{X}} \to \infty$, and weakly lower semicontinuous, if $\mathcal{R}(x) \leq \liminf_{n\to\infty} \mathcal{R}(x_n)$ for $(x_n)_{n\in\mathbb{N}} \rightharpoonup x$, where $\rightharpoonup$ denotes weak convergence, and $\to$ strong convergence. Any element in $\arg\min\{\mathcal{R}(x) \mid \mathbf{A}x = y\}$ is called an $\mathcal{R}$-minimizing solution of the equation $\mathbf{A}x = y$.

2.1 Convergence analysis

For networks $\mathbf{\Phi}, \mathbf{\Phi}_{\theta(\delta)}$ on $\mathbb{X}$ and $\delta > 0$ we define the noise-dependent regularizer $\mathcal{R}_\delta$ by (1.3), the limiting regularizer by $\mathcal{R}(x) = \|\mathbf{\Phi}(x) - x\|^2/2 + \lambda\, \mathcal{Q}(x)$ and consider noise-adaptive DD-Morozov regularization

$$\min_{x\in\mathbb{X}} \mathcal{R}_\delta(x) \quad \text{s.t.} \quad \|\mathbf{A}x - y_\delta\| \leq \delta\,. \tag{2.1}$$

Our results on the convergence of (1.3), (2.1) are derived under the following conditions, which we assume to be satisfied throughout this subsection.

Assumption 2.1.

(A1) $\mathbf{\Phi}, \mathbf{\Phi}_{\theta(\delta)}\colon \mathbb{X} \to \mathbb{X}$ *are weakly continuous.*

(A2) $\mathbf{\Phi}_{\theta(\delta)} \to \mathbf{\Phi}$ *weakly uniformly on bounded sets as* $\delta \to 0$.

(A3) $\mathbf{\Phi}_{\theta(\delta)} \to \mathbf{\Phi}$ *strongly pointwise on* $\mathcal{R}$*-minimizing solutions as* $\delta \to 0$.

(A4) $\mathcal{Q}\colon \mathbb{X} \to [0,\infty]$ *is proper, coercive and weakly lower semicontinuous.*

In (A2), weak uniform convergence on bounded sets means that for all bounded $B \subseteq \mathbb{X}$ and all $h \in \mathbb{X}$ we have $\sup_{x\in B}|\langle \mathbf{\Phi}_{\theta(\delta)}(x) - \mathbf{\Phi}(x), h\rangle| = 0$ as $\delta \to 0$. In the convergence analysis we assume that the networks $\mathbf{\Phi}_{\theta(\delta)}$ are trained, where $\theta(\delta)$ potentially depends on the noise, and all other quantities are given by the application or are user-specified. In many applications, the function $\mathbf{\Phi}_{\theta(\delta)}$ is a neural network for which $\|(\mathbf{\Phi}_{\theta(\delta)} - \mathrm{Id})(\cdot)\|$ may not be coercive which is the reason to add the term $\mathcal{Q}$ in (1.3).

Lemma 2.2. *The regularizers $\mathcal{R}, \mathcal{R}_\delta$ are coercive and weakly sequentially lower semicontinuous. Further, the feasible set $\{x \in \mathbb{X} \mid \|\mathbf{A}x - y_\delta\| \le \delta\}$ is weakly closed and non-empty for all $\delta > 0$ and all data y_δ with $\|\mathbf{A}x_\star - y_\delta\| \le \delta$ for some $x_\star \in \mathrm{dom}(\mathcal{Q})$.*

Proof. Let $(x_n)_{n\in\mathbb{N}} \in \mathbb{X}^{\mathbb{N}}$. Because $\mathcal{Q}$ is coercive and $\|\mathbf{\Phi}_{\theta(\delta)}(x) - x\|^2$ and $\|\mathbf{\Phi}(x) - x\|^2$ are non-negative, the functionals $\mathcal{R}, \mathcal{R}_\delta$ are coercive. Let $(x_n)_{n\in\mathbb{N}} \in \mathbb{X}^{\mathbb{N}}$ converge weakly to $x \in \mathbb{X}$. Because $\mathbf{\Phi}$ is weakly continuous, $(\mathbf{\Phi}(x_n) - x_n)_{n\in\mathbb{N}}$ converges weakly to $\mathbf{\Phi}(x) - x$. Due to the weak sequential lower semicontinuity of the norm, we infer $\|\mathbf{\Phi}(x)-x\|_{\mathbb{X}} \le \liminf_{n\to\infty} \|\mathbf{\Phi}(x_n)-x_n\|_{\mathbb{X}}$ which shows that $\mathcal{R}$ and in a similar manner $\mathcal{R}_\delta$ are weakly lower semicontinuous. Now, according to [5, Lemma 1.2.3], a functional $\mathcal{F}$ is weakly sequentially lower semicountinuous if and only if $\{x \in \mathbb{X} \mid \mathcal{F}(x) \le t\}$ is weakly sequentially closed for all $t > 0$. Because $\mathbf{A}$ is linear and bounded it is weakly continuous. Because the norm is weakly sequentially lower semicontinuous, $x \mapsto \|\mathbf{A}x - y_\delta\|$ is weakly sequentially lower semicontinuous, too. Hence $\{x \in \mathbb{X} \mid \|\mathbf{A}x - y_\delta\| \le \delta\}$ is weakly closed for all $\delta > 0$ and non-empty as it contains the exact data $\mathbf{A}x_\star$. □

Lemma 2.3 (Existence)**.** *For all data $y_\delta \in \mathbb{Y}$ with $\|\mathbf{A}x_\star - y_\delta\| \le \delta$ for some $x_\star \in \mathrm{dom}(\mathcal{Q})$, the constraint optimization problem* (2.1) *has at least one solution.*

Proof. Because $\mathcal{R}_\delta \ge 0$, the infimum M of $\mathcal{R}$ over $S_\delta := \{x \in \mathbb{X} \mid \|\mathbf{A}x - y_\delta\| \le \delta\}$ is nonnegative and there exists a sequence $(x_m)_m$ of elements of S_δ with $\lim_{m\to\infty} \mathcal{R}_\delta(x_m) = M$. Because $\mathcal{R}_\delta$ is coercive and $\{\mathcal{R}_\delta(x_m) \,|\, m \in \mathbb{N}\}$ is bounded, we infer that $(x_m)_m$ is bounded and thus there exists a weakly convergent subsequence $(x_{m(k)})_k$ converging to some $x \in \mathbb{X}$. Moreover, due to the weak closedness of S_δ we obtain $x \in S_\delta$. Because $\mathcal{R}_\delta$ is weakly lower semicontinuous, $\mathcal{R}(x) \le M$ and thus x is a solution of (1.2). □

Analogous to the proof of Lemma 2.3 one shows that there exists at least one $\mathcal{R}$-minimizing solution of $\mathbf{A}x = y$ whenever it is solvable in $\mathrm{dom}(\mathcal{Q})$.

Theorem 2.4 (Weak convergence)**.** *Let $\mathbf{A}x = y$ be solvable in $\mathrm{dom}(\mathcal{Q})$, $(y_n)_{n\in\mathbb{N}} \in \mathbb{Y}^{\mathbb{N}}$ satisfy $\|y - y_n\| \le \delta_n$, where $(\delta_n)_{n\in\mathbb{N}} \in (0,\infty)^{\mathbb{N}}$ with $\delta_n \to 0$, write $\mathcal{R}_n := \mathcal{R}_{\delta_n}$ and $\mathbf{\Phi}_n := \mathbf{\Phi}_{\theta(\delta_n)}$, and choose $x_n \in \arg\min\{\mathcal{R}_n(z) \mid \|\mathbf{A}z - y_n\| \le \delta_n\}$. Then $(x_n)_{n\in\mathbb{N}}$ has at least one weak accumulation point $x^\dagger \in \mathbb{X}$. Moreover, the limit of each weakly converging subsequence $(x_{n(k)})_{k\in\mathbb{N}}$ is an $\mathcal{R}$-minimizing solution of $\mathbf{A}x = y$ and $\mathcal{R}_{n(k)}(x_{n(k)}) \to \mathcal{R}(x^\dagger)$ for $k \to \infty$. If*

the $\mathcal{R}$-minimizing solution x^+ of $\mathbf{A}x = y$ is unique, then $x_n \rightharpoonup x^+$ and $\mathcal{R}_n(x_n) \to \mathcal{R}(x^+)$ as $n \to \infty$.

Proof. Set $S_n := \{z \in \mathbb{X} \mid \|\mathbf{A}z - y_n\| \leq \delta_n\}$ and $S_\star := \{z \in \mathbb{X} \mid \mathbf{A}z = y\}$. Clearly $S_\star \subseteq S_n$ and because $\mathbf{A}x = y$ is solvable, $S_\star$ is non-empty. Thus $\mathcal{R}_n(x_n) \leq \mathcal{R}_n(x^+) = \mathcal{Q}(x^+) + \|\mathbf{\Phi}_n(x_\star) - x_\star\|^2/2$, where $x_\star$ is an $\mathcal{R}$-minimizing solution of $\mathbf{A}x = y$. From the weak convergence of $\mathbf{\Phi}_n(x_\star)$ we see that the right hand side is bounded. Thus $\mathcal{Q}(x_n)$ is bounded and with the coercivity of $\mathcal{Q}$ we conclude there exists a weakly converging subsequence $(x_{n(k)})_{k\in\mathbb{N}}$. Because $\|\mathbf{A}(\cdot) - y\|$ is weakly lower semicontinuous,

$$\begin{aligned}\|\mathbf{A}(x^+) - y\| &\leq \liminf_{k\to\infty} \|\mathbf{A}(x_{n(k)}) - y\| \\ &\leq \liminf_{k\to\infty} \|\mathbf{A}(x_{n(k)}) - y_{n(k)}\| + \|y_{n(k)} - y\| \leq 2\delta_n \quad (2.2)\end{aligned}$$

and thus $x^+ \in S_\star$. It remains to verify that x^+ is an $\mathcal{R}$-minimizing solution of $\mathbf{A}x = y$. According to (A1), (A2) we have $\mathbf{\Phi}_{n(k)}(x_{n(k)}) \rightharpoonup \mathbf{\Phi}(x^+)$ and $\mathbf{\Phi}_n(x_\star) \to \mathbf{\Phi}(x_\star)$, and thus

$$\begin{aligned}\mathcal{Q}(x^+) + \frac{\lambda}{2}\|\mathbf{\Phi}(x^+) - x^+\| &\leq \liminf_{n\to\infty} \mathcal{Q}(x_n) + \frac{\lambda}{2}\|\mathbf{\Phi}_n(x_n) - x_n\| \\ &\leq \limsup_{n\to\infty} \mathcal{Q}(x_n) + \frac{\lambda}{2}\|\mathbf{\Phi}_n(x_n) - x_n\| \\ &\leq \limsup_{n\to\infty} \mathcal{Q}(x_\star) + \frac{\lambda}{2}\|\mathbf{\Phi}_n(x_\star) - x_\star\| \\ &= \mathcal{Q}(x_\star) + \frac{\lambda}{2}\|\mathbf{\Phi}(x_\star) - x_\star\|\,.\end{aligned}$$

Thus x^+ is an $\mathcal{R}$-minimizing solution of $\mathbf{A}x = y$ with $\mathcal{R}(x_{n(k)}) \to \mathcal{R}(x^+)$. Finally, if the $\mathcal{R}$-minimizing solution x^+ of $\mathbf{A}x = y$ is unique, then $(x_n)_{n\in\mathbb{N}}$ has exactly one weak accumulation point x^+ and $\mathcal{R}_n(x_n) \to \mathcal{R}(x^+)$. □

As in the paper [19], we introduce the concept of total nonlinearity, which is required for strong convergence.

Definiton 2.5 (Total nonlinearity)**.** *Let $\mathcal{F}\colon \mathbb{X} \to \mathbb{R}$ be Gâteaux differentiable at $x_\star \in \mathbb{X}$. The absolute Bregman distance $\mathcal{B}_\mathcal{F}(x_\star, \cdot)\colon \mathbb{X} \to [0, \infty]$ and modulus of total nonlinearity $\nu_\mathcal{F}(x_\star, \cdot) : (0, \infty) \to [0, \infty]$ of $\mathcal{F}$ at $x_\star$ are defined by*

$$\forall x \in \mathbb{X}\colon \quad \mathcal{B}_\mathcal{F}(x_\star, x) := |\mathcal{F}(x) - \mathcal{F}(x_\star) - \mathcal{F}'(x_\star)(x - x_\star)| \quad (2.3)$$

$$\forall t > 0\colon \quad \nu_\mathcal{F}(x_\star, t) := \inf\{\mathcal{B}_\mathcal{F}(x_\star, x) \,|\, x \in \mathbb{X} \wedge \|x - x_\star\|_\mathbb{X} = t\}\,. \quad (2.4)$$

The functional $\mathcal{F}$ is called totally nonlinear at $x_\star$, if $\nu_\mathcal{F}(x_\star, t) > 0$ for all $t \in (0, \infty)$.

According to [19], $\mathcal{F}$ is totally nonlinear at $x_\star \in \mathbb{X}$ if and only if for all bounded sequences

$(x_n)_{n\in\mathbb{N}} \in \mathbb{X}^{\mathbb{N}}$ with $\lim_{n\to\infty} \mathcal{B}_{\mathcal{F}}(x_\star, x_n) = 0$ we have $\lim_{n\to\infty} \|x_n - x_\star\|_{\mathbb{X}} = 0$.

Theorem 2.6 (Stong convergence)**.** *In the situation of Theorem 2.4 assume additionally that the $\mathcal{R}$-minimizing solution x^+ of $\mathbf{A}x = y$ is unique and that $\mathcal{R}$ is totally nonlinear at x^+. Then, $\|x_n - x^+\|_{\mathbb{X}} \to 0$ as $n \to \infty$.*

Proof. According to Theorem 2.4, the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to x^+ and $\mathcal{R}(x^+) = \lim_{n\to\infty} \mathcal{R}(x_n)$. Because $\mathcal{R}'(x^+)$ is bounded, $\mathcal{R}'(x^+)(x_n - x^+) \to 0$ and thus $\mathcal{B}_{\mathcal{R}}(x^+, x_n) = |\mathcal{R}(x_n) - \mathcal{R}(x^+) - \mathcal{R}'(x^+)(x_n - x^+)| \to 0$. Because $(x_n)_{n\in\mathbb{N}}$ is bounded with the total nonlinearity of $\mathcal{R}$ this yields $x_n \to x^+$. □

2.2 Training strategy

Given a sequence of noise levels $(\delta_n)_{n\in\mathbb{N}}$, our aim is to construct the data-driven regularizer $\|\mathbf{\Phi}_n(x) - x\|^2/2$ with neural networks $\mathbf{\Phi}_n$ adapted to training signals $x_i \in \mathbb{X}$ for $i \in I \coloneqq \{1, \dots, N\}$ that we consider as ground truth and corresponding perturbed signals $x_i + r_{i,j,n} \in \mathbb{X}$ for $j \in J_{i,n}$ that we want to avoid. Given a family $(\mathbf{\Phi}_\theta)_{\theta\in\Theta}$ we determine the parameter $\theta = \theta_n$ as the minimizer of

$$\mathcal{L}_n(\theta) \coloneqq \sum_{i\in I} \sum_{j\in J_{i,n}^\star} \|\mathbf{\Phi}(x_i + r_{i,j,n}) - x_i\|^2, \tag{2.5}$$

where $J_{i,n}^\star \coloneqq J_{i,n} \cup \{0\}$ and $r_{i,n,0} \coloneqq 0$ for the ground truth signals.

By doing so, we have $(\mathbf{\Phi} - \mathrm{Id})(x_i) \simeq 0$ for the ground truth signals x_i and $(\mathbf{\Phi} - \mathrm{Id})(x_i + r_{i,j,n}) \simeq r_{i,j,n}$ for the perturbed signals $x_i + r_{i,j,n}$. Hence the regularizer $\|(\mathbf{\Phi} - \mathrm{Id})(\cdot)\|$ is expected to be small for signals similar to x_i and large for signals similar to $x_i + r_{i,j,n}$. A specific feature of a learned regularizer is that it can depends on the forward problem. This is achieved by making the perturbations $r_{i,j,n}$ operator specific. A strategy for increasing this dependence is to let the architecture depend on $\mathbf{A}$ such as a null space network [30] or data-proximal network [10].

Remark 2.7 (Choice of the perturbations)**.** *A crucial question is how to construct proper perturbed signals $x_i + r_{i,j,n} \in \mathbb{X}$. For NETT, we proposed in [19] to choose a single perturbation $r_{i,1,n} = (\mathrm{Id} - \mathbf{A}^+\mathbf{A}) \in \ker(\mathbf{A})$ per training sample, independent of the noise. This choice is well suited to address non-uniqueness, which is the main problem in undersampled tomographic inverse problems where the kernel $\ker(\mathbf{A})$ is of high dimension. In this work, we are also interested in severely ill-posed problems where small singular values are a further main challenge. Therefore, we modify the training strategy of [19] by using multiple perturbations $r_{i,j,n}$ that take into account two additional issues: Some of the perturbations represent noise in the low-frequency components corresponding to large singular values, and some of them represent truncated high-frequency components of the signal corresponding to small singular values.*

Let $(u_n, v_n, \sigma_n)_{n\in\mathbb{N}}$ denote a singular value decomposition (SVD) of $\mathbf{A}$. Using the SVD, we can express $\mathbf{A}$, its pseudoinverse, and the truncated SVD reconstruction by the formulas

$$\begin{aligned}\mathbf{A}(x) &= \sum_{n\in\mathbb{N}} \sigma_n \langle x, u_n\rangle v_n \\ \mathbf{A}^{+}(y) &= \sum_{n\in\mathbb{N}} \sigma_n^{-1} \langle y, v_n\rangle u_n \\ \mathbf{S}_\alpha(y) &= \sum_{\sigma_n^2\geq\alpha} \sigma_n^{-1} \langle y, v_n\rangle u_n \,,\end{aligned}$$

where $\alpha \geq 0$ is the regularization parameter.

Now if x_i is a given ground truth signal and $y_{i,n} = \mathbf{A}x_i + z_{i,n}$ corresponding noisy data, and $\alpha[j,n]$ for $j \in J_n$ are variable chosen regularization parameters in truncated SVD, we consider perturbed signals

$$x_i + r_{i,j,n} \coloneqq \mathbf{S}_{\alpha[j,n]}(\mathbf{A}x_i + z_{i,n})\,.$$

In fact, the perturbed signals are truncated SVD regularized reconstructions with perturbations $r_{i,j,n} = \mathbf{S}_{\alpha[j,n]}(\mathbf{A}x_i + z_{i,j}) - x_i$. In particular, for $z_i = 0$ and $\alpha[j,n] = 0$, we recover the perturbations $\mathbf{A}^{+}\mathbf{A}x_i - x_i$ of [19], which are pure artifacts. The more general strategy that is proposed here also includes perturbations due to noise and to Gibbs-type artifacts caused by truncation of singular components.

3 Application

In this section we present numerical results for an ill posed problem related to attenuation correction in photoacoustic tomography (PAT). We consider discrete setting where the operator $\mathbf{A} \in \mathbb{R}^{d\times d}$ is a matrix of size $d = 600$ and $x \in \mathbb{R}^d$ is the time discretization of real valued function defined on the interval $[0, T]$. The additional regularizer $\mathcal{Q}$ is taken as total variation. Details on the forward operator, the learned regularizer and numerical solution of (2.1) are given below.

3.1 Implementation details

Forward operator: The forward operator $\mathbf{A}$ is taken as a discretization of a one-dimensional wave dissipation operator that models damping in PAT [18]. It corresponds to a discretized one-dimensional integral operator that maps unattenuated pressure signals to attenuated signals. Its operation is illustrated in Figure 3.1. For details on how the matrix $\mathbf{A}$ was simulated, see [13]. Due to the fast decay of the singular values of $\mathbf{A}$ (Figure 3.1 left), the solution of (1.1) is strongly ill-posed. Moreover, the operator is not of convolutional form, and the ill-

posedness increases for signal components corresponding to later times. This can be seen in the right image in Figure 3.1, where the right part of the signal is significantly more blurred. Also note the increased attenuation and shift to the right for later times.

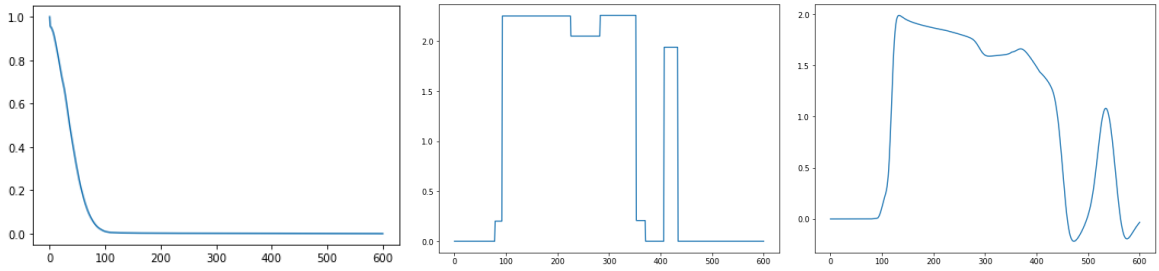

Figure 3.1: Left: Singular values of $\mathbf{A} \in \mathbb{R}^{d\times d}$ modeling dissipation. Middle: Test signal $x \in \mathbb{R}^d$. Right: Exact data $\mathbf{A}x$ for input from the middle picture. The horizontal axis in the middle and right images represents time.

Network architecture: The architecture $(\mathbf{\Phi}_\theta)_{\theta\in\theta}$ of the network resembles a one-dimensional version of the 2D Unet of [26] with one skip connection. It starts with two 1D convolutions of kernel size 7 and 16 channels, each followed by ReLU as the activation function. The spatial size of the feature maps is then halved using a max pooling layer. This convolution block is repeated twice, each time with twice as many filters (32 and 64). By applying three convolutional upsamplings, we obtain an output of the same spatial size as the input. We concatenate the input with the output and convolve one last time so that the network is able to learn the identity for the correct signals. We keep the network architecture quite simple to reduce the computational time of the gradient computation. Dropout layers are added to prevent overfitting.

Network training: For the training signals x_i we take collection of the block signal similar to the ones of [8]. We construct noisy data $\mathbf{A}x_i + z_{i,n}$ where $z_{i,n}$ is normally distributed noise with mean zero and standard deviation 0.1. The full training data set consists 5000 ground truth signals x_i and 5000 noisy signal for each $\alpha[j,n] = j/10$ with $j \in \{1,\dots,8\}$. According to Section 2.2, the network is trained by minimizing the risk $\mathcal{L}_n(\theta) = \sum_i \sum_j \|\mathbf{\Phi}_\theta(x_i + r_{i,j,n}) - x_i\|^2$. The trained network is then given by $\mathbf{\Phi}_n := \mathbf{\Phi}_{\theta_n}$, where θ_n is the numerical minimizer of $\mathcal{L}_n$.

Numerical DD-Morozov regularization: Reconstruction is done by numerically solving (2.1) with the noise-adaptive data-driven regularizer $\mathcal{R}_n := \|(\mathrm{Id} - \mathbf{\Phi}_n)(\cdot)\|^2/2 + \|\mathbf{L}x\|_1$, where $\|\cdot\|_1$ is the ℓ^1-norm and $\mathbf{L}$ is the discrete central difference operator with Neumann boundary

conditions. For that purpose, we write (2.1) in the form

$$\min_x \|\mathbf{\Phi}_n(x) - x\|_2^2 + \lambda\|\mathbf{L}x\|_1 + \mathbf{1}_B(\mathbf{A}x) \tag{3.1}$$

with $\mathbf{1}_B$ denoting the indicator function of $B = \{y \in \mathbb{R}^d \mid \|y - y_\delta\| \leq \delta\}$.

Optimization problem (3.1) is solved using the primal dual algorithm of [6]. With the abbreviations $F(x) := \|\mathbf{\Phi}_n(x) - x\|_2^2$, $h_1 := \lambda\|\cdot\|_1$ and $h_2 := \mathbf{1}_B$, parameters $\tau, \sigma, \rho > 0$ and initial values $x^{(0)}$ and $y^{(0)} = (0, 0)$ the proposed reconstruction algorithm reads

$$\begin{aligned}
z^{(i+1)} &:= x^{(i)} - \tau\nabla f(x^{(i)}) - \tau\mathbf{L}^* y_1^{(i)} - \tau\mathbf{A}^* y_2^{(i)} \\
x^{(i+1)} &:= \rho z^{(i+1)} + (1-\rho)x^{(i)} \\
w_1^{(i+1)} &:= \mathrm{prox}_{\sigma h_1^*}(y_1^{(i)} + \sigma\mathbf{L}(2z^{(i+1)} - x^{(i)})) \\
y_1^{(i+1)} &:= \rho w_1^{(i+1)} + (1-\rho)y_1^{(i)} \\
w_2^{(i+1)} &:= \mathrm{prox}_{\sigma h_2^*}(y_2^{(i)} + \sigma\mathbf{A}(2z^{(i+1)} - x^{(i)})) \\
y_2^{(i+1)} &:= \rho w_2^{(i+1)} + (1-\rho)y_2^{(i)}\,.
\end{aligned}$$

Here prox denotes the proximity mapping and $()^*$ the Fenchel dual. The proximity mappings $\mathrm{prox}_{\sigma h_1^*}$, $\mathrm{prox}_{\sigma h_2^*}$ can be easily computed using the relation $\mathrm{prox}_{h^*} + \mathrm{prox}_h = \mathrm{Id}$ and the known expressions for the proximity mappings of $\|\cdot\|_1$ and $\mathbf{1}_B$.

3.2 Results

Figure 3.2 shows results for a randomly selected block signal x that is not included in the training data.The upper left image shows the noisy attenuated signal y_δ, the upper right image shows the Backprojection (BP) reconstruction $\mathbf{A}^T y_\delta$, the lower left image shows the truncated SVD reconstruction $\mathbf{S}_\alpha y_\delta$ with $\alpha = 0.1$, and the lower right image shows the results with the proposed DD-Morozov regularization.The BP reconstruction is clearly damped, while the SVD reconstruction shows strong oscillations. The corresponding reconstruction using the DD-Morozov method (1.3), (2.1) is obviously much better. Similar results have been obtained for other randomly selected training signals.

Another example is shown in Figure 3.3, where we compare the ground truth, the network prediction of the truncated SVD, the Tikhonov regularization, and the DD-Morozov regularization. As regularization parameter α we have chosen the optimal value for the Tikhonov method. The network prediction of the truncated SVD is the trained network $\mathbf{\Phi}_n$ applied to the SVD reconstruction. We see that the Tikhonov regularization is worse than the network prediction of the truncated SVD, and the DD-Morozov regularization best recovers information about the original signal. This suggests that the regularization property of the Morozov

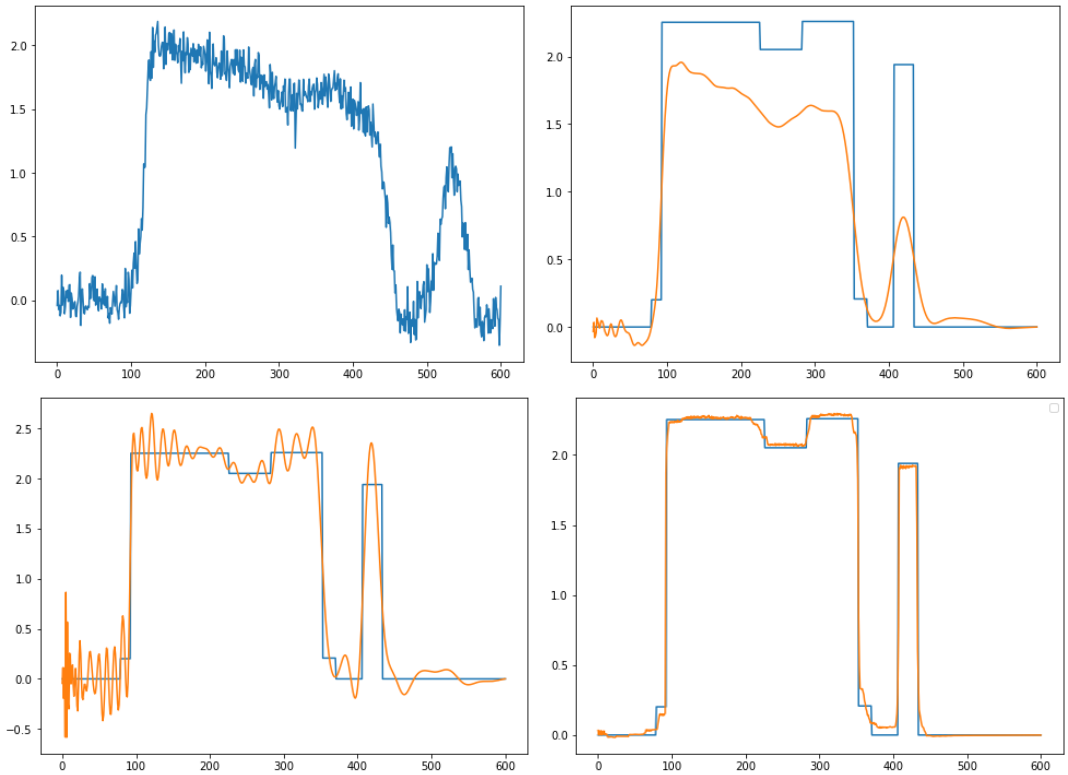

Figure 3.2: Top left: Noisy data y_δ. Top right: BP reconstruction. Bottom left: SVD reconstruction. Bottom right: DD-Morozov reconstruction.

method, together with a properly trained network, can be superior to either of these methods.

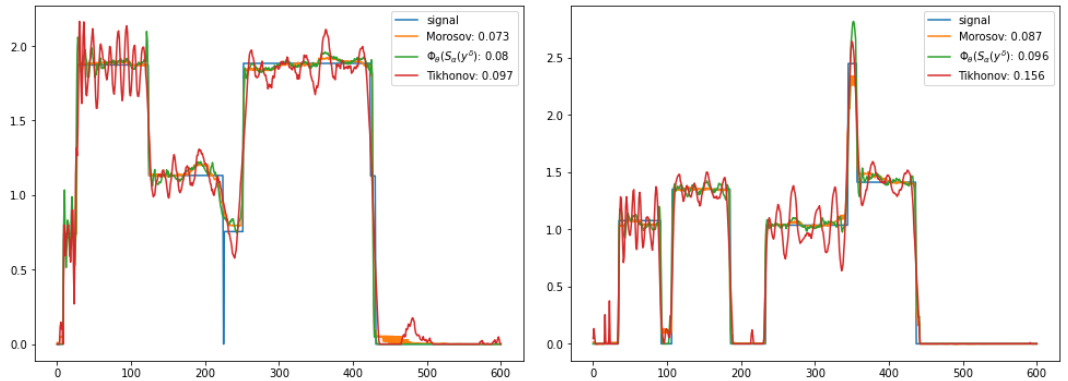

Figure 3.3: Two further randomly selected reconstruction results. In both figures, blue is the original signal, orange is the DD-Morozov reconstruction, green is the network prediction of truncated SVD, and red is the result of Tikhonov regularization.

4 Summary

In this paper, we introduced and analyzed neural network-based noise-adaptive Morozov regularization using a data-driven regularizer (NN-Morozov regularization). We performed a complete convergence analysis that also allows for noise-dependent regularizers. In addition, we established convergence in strong topology. To make our approach practical, we developed a simple yet efficient training strategy extending NETT [19]. We verified our methodology through numerical experiments, with a special focus on its application to attenuation correction for PAT. Our research can provide the basis for a broader integration of data-driven regularizers into various variational regularization techniques.

References

[1] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.

[2] A. Aspri, S. Banert, O. Öktem, and O. Scherzer. A data-driven iteratively regularized Landweber iteration. *Numerical Functional Analysis and Optimization*, 41(10):1190–1227, 2020.

[3] A. Aspri, Y. Korolev, and O. Scherzer. Data driven regularization by projection. *Inverse Problems*, 36(12):125009, 2020.

[4] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta numerica*, 27:1–111, 2018.

[5] P. Blanchard and E. Brüning. *Variational Methods in Mathematical Physics. A Unified Approach.* Springer-Verlag, Berlin, 1992.

[6] L. Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of optimization theory and applications*, 158(2):460–479, 2013.

[7] S. Dittmer, T. Kluth, P. Maass, and D. Otero Baguer. Regularization by architecture: A deep prior approach for inverse problems. *Journal of Mathematical Imaging and Vision*, 62:456–470, 2020.

[8] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[9] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[10] S. Göppel, J. Frikel, and M. Haltmeier. Data-proximal null-space networks for inverse problems. *arXiv:2309.06573*, 2023.

[11] A. Goujon, S. Neumayer, P. Bohra, S. Ducotterd, and M. Unser. A neural-network-based convex regularizer for inverse problems. *IEEE Transactions on Computational Imaging*, 2023.

[12] M. Grasmair, M. Haltmeier, and O. Scherzer. The residual method for regularizing ill-posed problems. *Applied Mathematics and Computation*, 218(6):2693–2710, 2011.

[13] M. Haltmeier, R. Kowar, and L. V. Nguyen. Iterative methods for pat in attenuating acoustic media. *Inverse Problems 33 (11)*, 2017.

[14] M. Haltmeier and L. Nguyen. Regularization of inverse problems by neural networks. In K. Chen, C.-B. Schönlieb, X.-C. Tai, and L. Younces, editors, *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pages 1–29. Springer International Publishing, Cham, 2021.

[15] K. Ito and B. Jin. *Inverse problems: Tikhonov theory and algorithms*, volume 22. World Scientific, 2014.

[16] V. K. Ivanov, V. V. Vasin, and V. P. Tanana. *Theory of linear ill-posed problems and its applications*, volume 36. Walter de Gruyter, 2013.

[17] E. Kobler, A. Effland, K. Kunisch, and T. Pock. Total deep variation for linear inverse problems. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 7549–7558, 2020.

[18] R. Kowar and O. Scherzer. Attenuation models in photoacoustics. In *Mathematical modeling in biomedical imaging II: Optical, ultrasound, and opto-acoustic tomographies*, pages 85–130. Springer, 2011.

[19] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier. NETT: Solving inverse problems with deep neural networks. *Inverse Problems*, 36(6):065005, 2020.

[20] S. Lunz, O. Öktem, and C.-B. Schönlieb. Adversarial regularizers in inverse problems. *Advances in neural information processing systems*, 31, 2018.

[21] V. A. Morozov. *Regularization Methods for Ill-Posed Problems*. CRC Press, Boca Raton, 1993.

[22] S. Mukherjee, M. Carioni, O. Öktem, and C.-B. Schönlieb. End-to-end reconstruction meets data-driven regularization for inverse problems. *Advances in neural information processing systems*, 34:21413–21425, 2021.

[23] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*, volume 5 of *Monographs on Mathematical Modeling and Computation*. SIAM, Philadelphia, PA, 2001.

[24] D. Obmann, L. Nguyen, J. Schwab, and M. Haltmeier. Augmented NETT regularization of inverse problems. *Journal of Physics Communications*, 5(10):105002, 2021.

[25] D. Riccio, M. J. Ehrhardt, and M. Benning. Regularization of inverse problems: Deep equilibrium models versus bilevel learning. *arXiv:2206.13193*, 2022.

[26] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[27] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 860–867. IEEE, 2005.

[28] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[29] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces*, volume 10. Walter de Gruyter, 2012.

[30] J. Schwab, S. Antholzer, and M. Haltmeier. Deep null space learning for inverse problems: convergence analysis and rates. *Inverse Problems*, 35(2):025008, 2019.

[31] A. N. Tikhonov and V. Arsenin. *Solutionsof ill-posed problems*. John Wiley & Sons, Washington, D.C., 1977.