This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

Stability of Image-Reconstruction Algorithms

Pol del Aguila Pla, *Member, IEEE,* Sebastian Neumayer, and Michael Unser, *Fellow, IEEE*

***Abstract*—Robustness and stability of image-reconstruction algorithms have recently come under scrutiny. Their importance to medical imaging cannot be overstated. We review the known results for the topical variational regularization strategies (ℓ_2 and ℓ_1 regularization) and present novel stability results for ℓ_p-regularized linear inverse problems for $p \in (1, \infty)$. Our results guarantee Lipschitz continuity for small p and Hölder continuity for larger p. They generalize well to the $L_p(\Omega)$ function spaces.**

***Index Terms*—Lipschitz continuity, Inverse problems, Variational problems, Bridge regression.**

I. Introduction

INVERSE problems are at the core of computational imaging. Medical imaging critically depends on the guarantees provided by established image-reconstruction methodologies to inform diagnostic and treatment decisions. New techniques based on artificial intelligence and deep neural networks offer major average performance improvements in most applications [1]–[4], at the cost of poor practical stability [5], [6] and a lack in theoretical guarantees. In particular, seemingly small perturbations of the measurements can produce large errors in the resulting image. Insidiously, these errors may incorporate deceptive patterns that look realistic because they were learnt from the training database (hallucination) [7]. Additionally, questions have also been raised on the stability guarantees provided by variational inverse-problem approaches using ℓ_p-regularization strategies to induce structure [8], [9]. In this paper, we first review the unicity and stability properties of classical Tikhonov regularization ($p = 2$) and sparsity-promoting regularization ($p = 1$). Then, we present novel stability results for ℓ_p-regularized inverse problems for $p \in (1, \infty)$. In particular, we show that the solution map is locally Lipschitz continuous for $p \in (1, 2]$ and globally $1/(p-1)$-Hölder continuous for $p \in (2, \infty)$. The proofs also cover the case of the $L_p(\Omega)$ function spaces. Our aim in presenting these results is to pave the way toward a quantitative comparison of image-reconstruction methods in terms of stability.

A broad category of image-reconstruction algorithms can be formulated as the variational problem [10]–[15]

$$\min_{\tilde{\mathbf{f}} \in \mathbb{R}^N} \left\{ \left\| \mathbf{y} - \tilde{\mathbf{A}}\tilde{\mathbf{f}} \right\|_2^2 + \lambda \left\| \mathbf{L}\tilde{\mathbf{f}} \right\|_p^p \right\}. \tag{1}$$

For $p = 2$, (1) corresponds to classical Tikhonov regularization [16], [17]; and for $p = 1$, (1) corresponds to sparsity-based regularization [18], [19]. Here, $\tilde{\mathbf{A}} \in \mathbb{R}^{M \times N}$ with $M \leq N$ is the forward operator. For a given imaging system, it relates the discrete image representation $\tilde{\mathbf{f}} \in \mathbb{R}^N$ to the measurements $\mathbf{y} \in \mathbb{R}^M$. Furthermore, $\mathbf{L}$ is a linear transform (*e.g.*, the finite-difference operator) that gets penalized through the ℓ_p norm, and $\lambda \in \mathbb{R}_+$ is the regularization parameter controlling the tradeoff between the data-fidelity term and the regularizer. An alternative formulation to (1) is the synthesis formulation

$$\min_{\mathbf{f} \in \mathbb{R}^N} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{f}\|_2^2 + \lambda \|\mathbf{f}\|_p^p \right\}, \tag{2}$$

which, if $\mathbf{L}$ is invertible, corresponds exactly to (1) with $\mathbf{f} = \mathbf{L}\tilde{\mathbf{f}}$ and $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{L}^{-1}$. Beyond image reconstruction (*e.g.*, for template-based reconstruction methods [20]–[23]), this type of variational problem appears in, for example, statistics under the name of bridge regression [24], [25] for $p \in (1, 2)$, and in machine learning as part of the multiple-kernel learning [26] literature.

The objective of this paper is to study the robustness of the reconstruction of $\mathbf{f}$ from $\mathbf{y}$ based on (2). Although concrete definitions of robustness, stability, and similar concepts vary, the predominant view in the literature is that robustness should be measured in terms of the continuity properties of the reconstruction (or solution) map $S \colon \mathbb{R}^M \to \mathbb{R}^N$. This map is only well defined if (2) has a unique solution $\mathbf{f}_\mathbf{y}$, in which case $S \colon \mathbf{y} \mapsto \mathbf{f}_\mathbf{y}$. In this context, we study the stability of the reconstruction in terms of bounds on $\|\mathbf{f}_{\mathbf{y}_1} - \mathbf{f}_{\mathbf{y}_2}\|_{\ell_p}$ with respect to $\|\mathbf{y}_1 - \mathbf{y}_2\|_2$ for any two measurements $\mathbf{y}_1$ and $\mathbf{y}_2$. Depending on the relation between these terms, the stability is weaker or stronger. The most general category we contemplate for stability is local Hölder continuity, where

$$\|\mathbf{f}_{\mathbf{y}_1} - \mathbf{f}_{\mathbf{y}_2}\|_{\ell_p} \leq K \|\mathbf{y}_1 - \mathbf{y}_2\|_2^\beta, \tag{3}$$

with $K \geq 0$ and $\beta \in [0, 1]$ for any two measurements $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$ within a set of measurements $Y \subset \mathbb{R}^M$. Here, local signifies that $K = K(Y)$ depends on the choice of set Y, which may be a cube or ball in $\mathbb{R}^M$ containing expected reasonable measurements. The strongest stability result comprised within the same expression (3) is global Lipschitz continuity, where $Y = \mathbb{R}^M$ and $\beta = 1$. Given bounds such as (3) for any two image-reconstruction algorithms, one can objectively compare their stability properties in terms of the exponent β and the value of K.

A. Related Work

Although the robustness of regularized variational problems has been studied extensively before, most studies relied on asymptotic criteria for vanishing noise [27]–[30]. These are valid only when $\|\mathbf{y}_1 - \mathbf{y}_2\|_2 \to 0$ and are weaker than the ones we target under the conditions stated in (3). (See [31] for an extensive overview on stability criteria for variational problems.)

The stability of solutions of variational inverse problems has also been investigated using criteria similar to ours. In [32],

Pol del Aguila Pla is with the CIBM Center for Biomedical Imaging, in Switzerland.

The authors are with the Biomedical Imaging Group at the École polytechnique fédérale de Lausanne, in Lausanne, Switzerland.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

the authors assume that the forward operator is injective and invertible and that directional derivatives do not vanish at non-smooth points of the objective functional. These conditions are rather restrictive and superfluous in our particular setting. In [33], the authors consider finite-dimensional constrained-optimization problems and use ideas similar to ours. However, our analysis builds on a condition that involves the modulus of convexity of the regularizer and that is less limiting than the strong convexity imposed in [33].

A related but fundamentally different problem than the one we discuss is algorithmic stability in learning theory. There, the interest is to bound the magnitude of changes in the output of a learned algorithm with respect to changes in its training data. In that context, ℓ_1 and ℓ_p regularization have also been studied in detail [34].

II. Variational Regularization of Inverse Problems

We now discuss the variational regularization of linear inverse problems from the perspective used in [35]–[39]. The theory is formulated for Banach spaces, which are complete vector spaces with a norm. This level of generality is appropriate for our study because the ℓ_p spaces that characterize (2) are Banach spaces. Throughout the main body of the paper, we rely on the intuitive understanding of some of the mathematical terms, without diverting the reader's attention with extensive technical details. Appendix A is designed to complement the paper by providing the basic functional analytical background for our work.

An image f is considered as an object in a Banach space $\mathcal{X}$. The measurements $\mathbf{y} \in \mathbb{R}^M$ of f are modeled as some noisy version of $\boldsymbol{\nu}(f)$, where $\boldsymbol{\nu}\colon \mathcal{X} \to \mathbb{R}^M$ is a linear operator given by

$$f \mapsto \boldsymbol{\nu}(f) \coloneqq \big(\langle \nu_1, f\rangle_{\mathcal{X}'\times\mathcal{X}}, \ldots, \langle \nu_M, f\rangle_{\mathcal{X}'\times\mathcal{X}}\big), \tag{4}$$

for the set $\{\nu_m\}_{m=1}^M \subset \mathcal{X}'$ of linearly independent measurement functionals. Here, the ν_m are elements of the dual space $\mathcal{X}'$, which is made of all the linear continuous functionals $\nu\colon \mathcal{X} \to \mathbb{R}$ (see Definition 15). The notation $\langle \nu_m, f\rangle_{\mathcal{X}'\times\mathcal{X}}$ is used to denote the evaluation of ν_m at f. The operator $\boldsymbol{\nu}$ in (4) generalizes the role of the matrix $\mathbf{A}$ in (2). The choice of the pair $(\mathcal{X}, \mathcal{X}')$ for a specific problem corresponds to the choice of regularizer and, thus, to the choice of the desired properties of f (see (5) below). Although the theory is more general [39] we make here the restrictive assumption that $\mathcal{X}$ is a reflexive and strictly convex Banach space (see Definition 21). This is true for the spaces of interest in this paper, which are $\mathcal{X} = \ell_p$ and $\mathcal{X}' = \ell_q$ with $p \in (1, \infty)$ and $q = p/(p-1)$. Then, the solutions to the variational problem

$$\min_{f\in\mathcal{X}} \{E(\mathbf{y}, \boldsymbol{\nu}(f)) + \psi(\|f\|_{\mathcal{X}})\} \tag{5}$$

are taken to be reconstructions of f from the measurements $\mathbf{y}$.

The function $E\colon \mathbb{R}^M \times \mathbb{R}^M \to \overline{\mathbb{R}}_+$ is a data-fidelity term. It penalizes reconstructions f that do not agree with the measurements $\mathbf{y}$; for example, this could be the least-squares term used in (2). The function E is assumed to be lower semi-continuous, proper, and strictly convex in its second argument. The function $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be strictly increasing and strictly convex. It regulates how much we penalize f according to its norm $\|f\|_{\mathcal{X}}$. In general, more than one reconstruction f may achieve the same minimum cost

$$J(\mathbf{y}, f) = E(\mathbf{y}, \boldsymbol{\nu}(f)) + \psi(\|f\|_{\mathcal{X}}). \tag{6}$$

Such reconstructions are considered as equally good for the variational problem (5). As an example of the setup above, in the case of two-dimensional computed tomography (CT) problems, one can usually model X-Ray detectors using an impulse response $h \in L_q(\mathbb{R})$ for some $q \in (1, \infty)$, so that the M measurements at angles $\{\theta_m\}_{m=1}^M \subset [0, \pi)$ and offsets $\{t_m\}_{m=1}^M \subset \mathbb{R}$ are given by $\nu_m = h(t_m - \langle \cdot, \boldsymbol{\theta}_m\rangle)$, where $\boldsymbol{\theta}_m = [\cos(\theta_m), \sin(\theta_m)]$. Then, we have that $\nu_m \in L_q(\Omega) = \mathcal{X}'$ for any bounded $\Omega \subset \mathbb{R}^2$, and thus $\mathcal{X} = L_p(\Omega)$ with $p = 1/(1-1/q)$. Then, the regularizer in (5) can be chosen as the L_p regularizer.

The main object of study in this paper, the optimization problem (2), corresponds to the setup above with

$$\mathcal{X} = (\mathbb{R}^N, \|\cdot\|_p),\ \psi(x) = \lambda x^p, \text{ and } \lambda \in \mathbb{R}_+. \tag{7}$$

Theorem 1 characterizes the solutions of (5) in full generality using the duality map of the dual space $\mathcal{X}'$. In this setting, the duality map is the nonlinear map $\mathcal{J}_{\mathcal{X}'}\colon \mathcal{X}' \to \mathcal{X}$ that generalizes the concept of parallel vectors to Banach spaces (see Proposition 17 and Definition 18).

Theorem 1 (Representer Theorem for Inverse Problems [39])**.** *For a reflexive and strictly convex pair $(\mathcal{X}, \mathcal{X}')$ of Banach spaces with duality map $\mathcal{J}_{\mathcal{X}'}\colon \mathcal{X}' \to \mathcal{X}$, the variational problem (5) has the unique solution*

$$f_{\mathbf{y}} = \mathcal{J}_{\mathcal{X}'}(\nu_{\mathbf{y}}), \text{ with } \nu_{\mathbf{y}} = \sum_{m=1}^{M} \mathbf{a}_{\mathbf{y}}[m]\nu_m, \tag{8}$$

for a unique coefficient vector $\mathbf{a}_{\mathbf{y}} \in \mathbb{R}^M$.

Theorem 1 reveals several favorable properties of the variational approach to inverse problems. First, it guarantees that

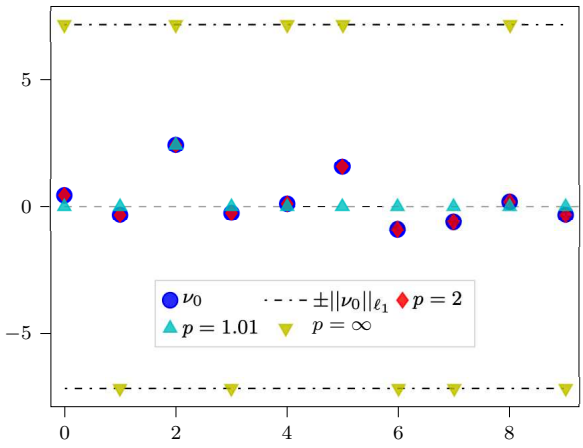

Fig. 1. Application of different duality maps $\mathcal{J}_{\mathcal{X}'}$ to a vector $\nu_0 \in \mathcal{X}'$ when $\mathcal{X} = (\mathbb{R}^{10}, \|\cdot\|_p) \subset \ell_p$ for different values of p.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

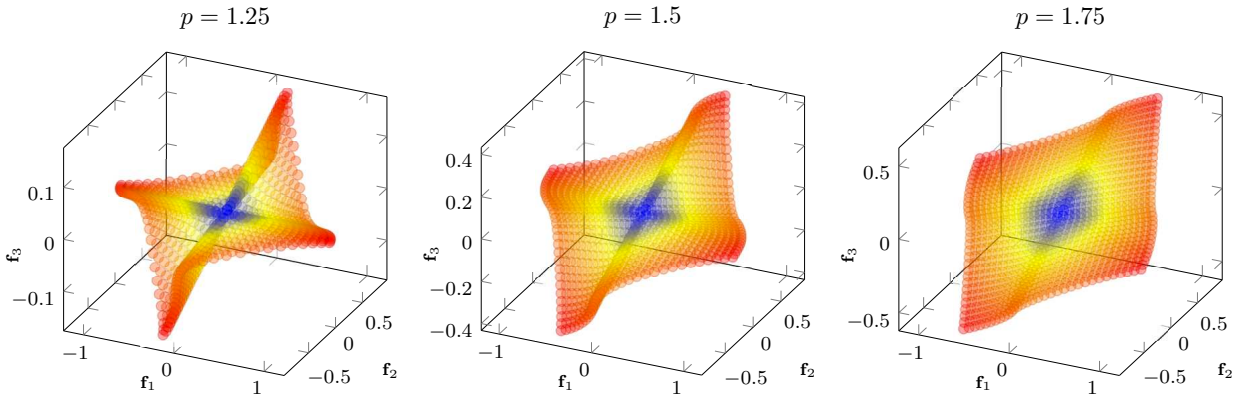

Fig. 2. Dependence of $\mathbf{f} = \mathcal{J}_{\mathcal{X}'}(\nu)$ on the vector of coefficients $\mathbf{a}$, with $\nu = \mathbf{a}[1]\nu_1 + \mathbf{a}[2]\nu_2$. The norm of the vector of coefficients is encoded in the intensities. Here, $\mathcal{X} = (\mathbb{R}^3, \|\cdot\|_p)$ for $p \in \{1.25, 1.5, 1.75\}$, the measurement operator $\boldsymbol{\nu} : \mathbb{R}^3 \to \mathbb{R}^2$ is composed of two ℓ_2-normalized random measurement vectors ν_1 and ν_2, and a regular (32×32) grid of $\mathbf{a}$ in $[-1, 1]^2$ is explored.

(5) always has a well-defined, unique solution. Further, (8) transforms the search for $f_{\mathbf{y}} \in \mathcal{X}$ (in the case of (2), of dimension $N \geq M$ and, in general, possibly infinite-dimensional) into a finite-dimensional search for $\mathbf{a}_{\mathbf{y}} \in \mathbb{R}^M$, a vector of the same dimension as the measurements. Prior information is injected in the solution by means of the duality map $\mathcal{J}_{\mathcal{X}'}$ given by the chosen regularization—through the choice $(\mathcal{X}, \mathcal{X}')$—which maps $\nu_{\mathbf{y}} \mapsto f_{\mathbf{y}}$. The regularizing effect of the duality map can be seen in Figure 1. The general result of Theorem 1 fits within a family of representer theorems for variational problems [35], [37]–[44].

For the variational problem (2), Theorem 1 guarantees that the reconstruction map $S \colon \mathbb{R}^M \to \mathcal{X}$ is well defined and that it is the composition of an unknown map $\mathbf{y} \mapsto \nu_{\mathbf{y}}$ with the duality map $\mathcal{J}_{\mathcal{X}'}$. Thereby, the study of $\mathcal{J}_{\mathcal{X}'}$ offers a first, more intuitive, approximation to the problem. Fortunately, its expression is known in closed-form for $\mathcal{X} = \ell_p$ (see Proposition 19). Figure 2 illustrates the duality map with small computational examples for $p \in \{1.25, 1.5, 1.75\}$, where we control the coefficient vector $\mathbf{a}_{\mathbf{y}} \in \mathbb{R}^2$ and depict the resulting solutions $\mathbf{f}_{\mathbf{y}} \in \mathbb{R}^3$. These suggest that stability highly depends on p, with rougher landscapes as p approaches 1 and smoother ones as it approaches 2.

As a prelude to Section III, where we characterize the robustness of image reconstruction using ℓ_p regularization for $p \in (1, \infty)$, we now discuss the two topical examples of ℓ_2 regularization and sparsity-promoting ℓ_1 regularization. Although known for the most part, the results will set the context for the later, more general results for $p \in (1, \infty)$.

A. ℓ_2 Regularization and Hilbert Spaces

The classical example of variational regularization is ℓ_2 regularization, as in (7) with $p = 2$. As it turns out, the analysis of the stability of the solution is not specific to the finite-dimensional case where $\mathbf{f} \in \mathbb{R}^N$. It can be transposed into the more general space of finite-energy discrete signals $f \in \ell_2$. In fact, because ℓ_2 is a Hilbert space (a Banach space with an inner product, see Definition 14), ℓ_2 regularization can be analyzed in the broader setting of Tikhonov regularization in Hilbert spaces. The following analysis is valid for every Hilbert space $\mathcal{H}$, including $L_2(\Omega)$ and other Hilbert/Sobolev function spaces. In the context of Theorem 1, then, this corresponds to choosing $\mathcal{X} = \mathcal{H}$. Because Hilbert spaces are strictly convex and reflexive, Theorem 1 applies. The duality map in Hilbert spaces with the choices above corresponds to the Riesz map $\mathcal{R} \colon \mathcal{H}' \to \mathcal{H}$ (see Definition 20 and the subsequent discussion). The Riesz map is linear. Thus, it holds that

$$f_{\mathbf{y}} = \sum_{m=1}^{M} \mathbf{a}_{\mathbf{y}}[m] \varphi_m, \text{ where } \varphi_m = \mathcal{R}\{\nu_m\} \tag{9}$$

for a unique vector of coefficients $\mathbf{a}_{\mathbf{y}} \in \mathbb{R}^M$. Using that $\|f\|_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}}$ for any $f \in \mathcal{H}$ (see Definition 14), and that $\langle \nu_m, \varphi_n \rangle_{\mathcal{H}' \times \mathcal{H}} = \langle \varphi_m, \varphi_n \rangle_{\mathcal{H}}$ (see Definition 20), where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product in $\mathcal{H}$, we obtain that

$$\|f_{\mathbf{y}}\|_{\mathcal{H}}^2 = \mathbf{a}_{\mathbf{y}}^{\mathrm{T}} \mathbf{H} \mathbf{a}_{\mathbf{y}}, \text{ and } \boldsymbol{\nu}(f_{\mathbf{y}}) = \mathbf{H} \mathbf{a}_{\mathbf{y}}\,, \tag{10}$$

with

$$\mathbf{H} = \left(\langle \varphi_m, \varphi_n \rangle_{\mathcal{H}} \right)_{n,m \in \{1,\ldots,M\}} \in \mathbb{R}^{M \times M}.$$

Using (9) and (10), we then write (5) as the finite-dimensional optimization problem

$$\min_{\mathbf{a} \in \mathbb{R}^M} \left\{ E\left(\mathbf{y}, \mathbf{H}\mathbf{a}\right) + \lambda\, \mathbf{a}^{\mathrm{T}} \mathbf{H} \mathbf{a} \right\}. \tag{11}$$

For the specific case of least squares, where E is chosen as

$$E(\mathbf{y}_1, \mathbf{y}_2) = \frac{1}{2} \|\mathbf{y}_1 - \mathbf{y}_2\|_2^2, \tag{12}$$

this results in a fully quadratic problem on the coefficients $\mathbf{a}$, with closed-form solution

$$\mathbf{a}_{\mathbf{y}} = \left(\mathbf{H}^{\mathrm{T}} \mathbf{H} + 2\lambda \mathbf{H}^{\mathrm{T}} \right)^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{y} = \left(\mathbf{H} + 2\lambda \mathbf{Id} \right)^{-1} \mathbf{y}, \tag{13}$$

where we took advantage of the property that $\mathbf{H}$ is Hermitian by construction and full-rank due to the linear independence of the measurement functionals ν_m. This allows us to characterize stability as in Proposition 2.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

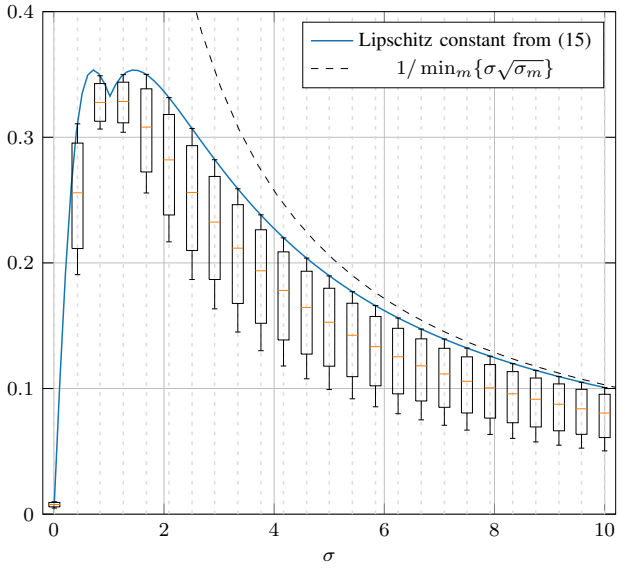

Fig. 3. Lipschitz constant in (15) compared to experimental values of $\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_2/\|\mathbf{y}_1 - \mathbf{y}_2\|_2$ for ℓ_2-regularized least-squares problems with $\lambda = 1$ and $\tilde{\boldsymbol{\nu}} = \sigma\boldsymbol{\nu}$, where $\boldsymbol{\nu}\colon \mathbb{R}^3 \to \mathbb{R}^2$ has random-normal entries.

Proposition 2 (Lipschitz Continuity of Tikhonov-Regularized Least Squares)**.** *Consider the ℓ_2-regularized least-squares optimization problem*

$$\min_{f\in\ell_2}\Big\{\frac{1}{2}\,\|\mathbf{y} - \boldsymbol{\nu}(f)\|_2^2 + \lambda\,\|f\|_{\mathcal{H}}^2\Big\}. \tag{14}$$

Moreover, consider two measurements $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$ and their associated solutions of (14), $f_{\mathbf{y}_1}, f_{\mathbf{y}_2} \in \ell_2$. Then, one has that

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\mathcal{H}} \leq \max_{m\in\{1,2,\ldots,M\}} \frac{\sqrt{\sigma_m}}{\sigma_m + 2\lambda}\|\mathbf{y}_1 - \mathbf{y}_2\|_2 \tag{15}$$

and the reconstruction map $S\colon \mathbf{y} \mapsto f_{\mathbf{y}}$ is Lipschitz continuous. Here, $\{\sigma_m\}_{m=1}^M$ is the set of eigenvalues of the matrix $\mathbf{H}$.

Proof. Consider (13) and (10). Then,

$$\begin{aligned}\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\mathcal{H}}^2 &= (\mathbf{f}_{\mathbf{y}_1} - \mathbf{f}_{\mathbf{y}_2})^T\mathbf{H}\,(\mathbf{f}_{\mathbf{y}_1} - \mathbf{f}_{\mathbf{y}_2}) \\ &= (\mathbf{y}_1 - \mathbf{y}_2)^{\mathrm{T}}\mathbf{P}\frac{\boldsymbol{\Lambda}}{(\boldsymbol{\Lambda} + 2\lambda\mathbf{Id})^2}\mathbf{P}^T(\mathbf{y}_1 - \mathbf{y}_2) \\ &\leq \max_{m\in\{1,\ldots,M\}} \frac{\sigma_m}{(\sigma_m + 2\lambda)^2}\|\mathbf{y}_1 - \mathbf{y}_2\|_2^2, \quad (15')\end{aligned}$$

where $\mathbf{P}$ is the orthogonal matrix of eigenvectors of $\mathbf{H}$ such that $\mathbf{H} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix containing the eigenvalues of $\mathbf{H}$. □

Of particular interest is how the bound (15) scales with respect to the measurement functionals ν_m and their Gram matrix $\mathbf{H}$ for a given regularization parameter λ. For instance, if we consider a measurement operator $\tilde{\boldsymbol{\nu}} = \sigma\boldsymbol{\nu}$, we see that the Lipschitz constant in (15) decays as $1/\min_m\{\sigma\sqrt{\sigma_m}\}$ for $\sigma \to \infty$. In other words, stability is ultimately regulated by those changes to which $\tilde{\boldsymbol{\nu}}$ is least sensitive. The behavior of the bound with respect to σ as well as empirical results are portrayed in Figure 3, together with the expected asymptotic behavior.

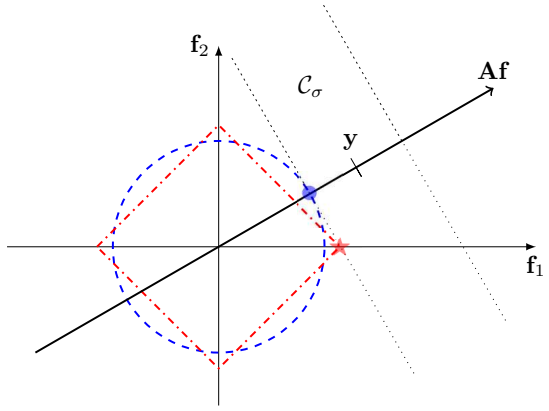

Fig. 4. Pictorial representation of the solution to $\min_{\mathbf{f}\in\mathbb{R}^N}\|\mathbf{f}\|_p$ subject to $\|\mathbf{Af} - \mathbf{y}\|_2 \leq \sigma$ for $p \in \{1, 2\}$. Here, $\mathbf{A} \in \mathbb{R}^{1\times 2}$ and $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2) \in \mathbb{R}^2$. The solution is unique both for $p = 1$ and $p = 2$, while the ℓ_1 penalty leads to the sparse solution, indicated by a star.

As expected, in (15) we see that increasing the regularization parameter λ will result in more stable solution maps. However, doing so will also increase the bias of the resulting algorithm—c.f. (13)—negatively affecting performance. Our results throughout the paper aim to compare the stability of algorithms once all parameters have been selected to obtain the best achievable performance.

To summarize, ℓ_2 regularization (and, in general, Tikhonov regularization in any Hilbert space) leads to a unique solution, with a solution map that is globally Lipschitz continuous. Although Proposition 2 only covers least-squares problems, we shall see in Section III that this remains true for any other strictly convex data-fidelity term.

B. ℓ_1 Regularization and Sparsity

Variational image reconstruction driven by sparsity-promoting regularization using the ℓ_1 norm is supported by the theory known as compressed sensing [18], [19]. The optimization problem

$$\min_{\mathbf{f}\in\mathbb{R}^N}\|\mathbf{f}\|_1 \text{ subject to } \|\mathbf{Af} - \mathbf{y}\|_2 \leq \sigma \tag{16}$$

has for solution the sparsest vector within the constraint set $\mathcal{C}_\sigma = \{\mathbf{f} \in \mathbb{R}^N : \|\mathbf{Af} - \mathbf{y}\|_2 \leq \sigma\}$, provided $\mathbf{A}$ fulfills some rather strict conditions, namely, the restricted-isometry property. The minimization in (16) is portrayed in Figure 4, where the radii of the ℓ_2- and ℓ_1-norm balls are increased until they meet the boundary of $\mathcal{C}_\sigma$. In that example, the ℓ_1-norm minimization does indeed lead to a sparse solution. Although this analysis relies on the constrained formulation in (16), the effect of the ℓ_1-norm on the set of solutions of the corresponding regularized formulation is effectively the same [15, Remark 3.3]. This regularized formulation corresponds to (7) when $p = 1$.

The same restricted-isometry property that guarantees a unique, sparse solution to (16) also provides the Lipschitz stability [18] of that solution with respect to variations in the measurements. In particular, when $\mathbf{A}$ is composed of linearly independent measurement functionals, as assumed in

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

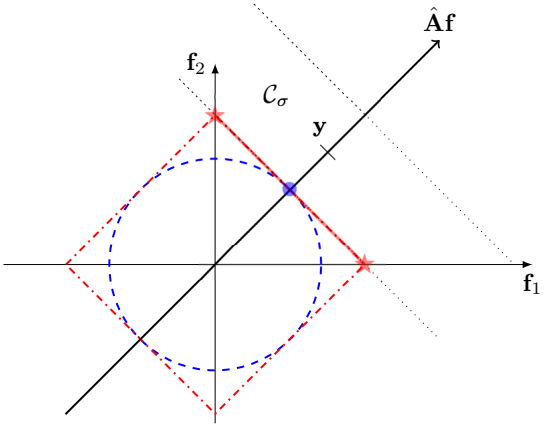

Fig. 5. Same representation as in Figure 4, but for another linear operator. Here, the ℓ_1 penalty leads to a problem with infinitely many solutions. In particular, the solution set contains an entire edge of the ℓ_1-norm unit ball: the convex hull of $\{(0,1),(1,0)\}$. This generalizes well to the infinite-dimensional sparsity-promoting regularization setting, see [35].

Theorem 1, it can be shown that there exists a constant $K_{\mathrm{s}} \in \mathbb{R}_+$ such that

$$\|\mathbf{f}_{\mathbf{y}_1} - \mathbf{f}_{\mathbf{y}_2}\|_2 \leq K_{\mathrm{s}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2. \tag{17}$$

The caveat, however, is that it is challenging (or impossible) to obtain or bound the constant K_{s} for specific practical problems. Because compressed sensing is formulated for finite-dimensional spaces, norm equivalence, together with (17), imply Lipschitz stability, as initially described in (3) ($\beta = 1$ and $Y = \mathbb{R}^M$). In particular, the fact that $\|\mathbf{f}\|_1 \leq K_{1,2}\|\mathbf{f}\|_2$ for any $\mathbf{f} \in \mathbb{R}^N$ and some fixed $K_{1,2} \in \mathbb{R}_+$ implies that $K = K_{1,2}K_{\mathrm{s}}$.

These attractive theoretical results, however, cannot be applied to many sparsity-based image-reconstruction algorithms, particularly to those that fail to operate within the regime of the restricted-isometry property. When this happens, one may be faced with an infinite number of solutions of (16), as portrayed in Figure 5. Then, the corresponding regularized problem has to be characterized by a version of the representer theorem that is more general than the one we presented in Theorem 1 [39, Section 4]. Indeed, the extended version contemplates Banach spaces that are not strictly convex and the solution set is shown to be the convex hull of a number of sparse extreme points. Note that, under slightly more general conditions than the restricted isometry property, weaker forms of stability than the ones we are interested in can still be obtained for ℓ_1 regularization [45], [46].

III. Stability of Solutions

In this section, we present our results for the general case of ℓ_p regularization with $p \in (1,\infty)$. Because, as in the case of ℓ_2 regularization in Section II-A, the analysis of the stability of the solution is not particular to the finite-dimensional setting, we directly expose it for a (potentially infinitely-supported) discrete signal $f \in \ell_p$. We study the problem

$$\min_{f \in \ell_p} \left\{ E\left(\mathbf{y}, \boldsymbol{\nu}(f)\right) + \lambda \|f\|_{\ell_p}^p \right\} \tag{18}$$

TABLE I
Continuity bounds obtained on the stability of solutions to ℓ_p-regularized inverse problems.

$p \in (1,2)$	$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_p} \leq \dfrac{(2\,r_p(Y))^{2-p} K_p}{\lambda p(p-1)} \|\mathbf{y}_1 - \mathbf{y}_2\|_2$
$p \in [2,\infty)$	$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_p} \leq \left(\dfrac{2^{p-2}K_p}{\lambda p}\right)^{\frac{1}{p-1}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2^{\frac{1}{p-1}}$

with $\lambda \in \mathbb{R}_+$, under the following assumptions:

(A1) the data-fidelity term $E\colon \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_+$ is lower-semicontinuous and strictly convex in its second argument;

(A2) the data-fidelity term E is differentiable in its second argument, and the gradient $\nabla_f \{E(\mathbf{y}, \boldsymbol{\nu}(f))\}$ is Lipschitz continuous with respect to changes in the measurements $\mathbf{y}$ with constant K_p for any f;

(A3) the measurement operator $\boldsymbol{\nu}$ is composed of M linearly independent functionals $\{\nu_m\}_{m=1}^M$ as in (4).

We now discuss our main results (Theorems 3 and 4). The proofs are deferred to their more general cases (Theorems 11 and 12), which cover the case of $L_p(\Omega)$ function spaces.

A. Hölder Continuity for $p \in [2,\infty)$

The case of $p = 2$ has been discussed for least-squares problems in Section II-A. There, we proved Lipschitz continuity of the solution using a Hilbert-space analysis (see Proposition 2). Our result here extends this claim for $p = 2$ to any E that fulfills (A1) and (A2).

Theorem 3 (Hölder Continuity of ℓ_p-Regularized Linear Inverse Problems, $p \in [2,\infty)$)**.** *Consider the variational problem (18) with $p \in [2,\infty)$ and assume that (A1)–(A3) are fulfilled. Then, for any given measurement $\mathbf{y} \in \mathbb{R}^M$, (18) has a unique solution $f_{\mathbf{y}} \in \ell_p$. Moreover, it holds that*

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_p} \leq \left(\frac{2^{p-2}K_p}{\lambda p}\right)^{\frac{1}{p-1}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2^{\frac{1}{p-1}}, \tag{19}$$

for any two $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$.

The exponent $\beta = 1/(p-1)$ in (19) characterizes the continuity bound of Theorem 3. In particular, this exponent takes unit value for $p = 2$, which indeed leads to Lipschitz continuity, *cf.* (3) with $\beta = 1$ and $Y = \mathbb{R}^M$. For $p > 2$, the result is weaker because, in the low-noise regime where $\|\mathbf{y}_1 - \mathbf{y}_2\|_2 \to 0$, the bound on $\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_p}$ gets larger as p increases. The change of the type of continuity result for different values of p is an unexpected result. One might think that this is caused by the mismatch between the norms used in either side of (19). However, because all norms are equivalent in the finite-dimensional space of measurements $\mathbb{R}^M$, the exponent in (19) holds for every other norm in $\mathbb{R}^M$,

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

and the choice of any other specific norm would simply be absorbed by the preceding constant.

The term K_p that appears in the leading constant of the bound characterizes the role of the data-fidelity term E on the stability of the solution. Most importantly, and as opposed to what happens in (17) with compressed sensing, all terms in the bound of Theorem 3 can be computed or bounded in practical applications. In particular we evaluate our result in Section III-C in the case of ℓ_p-regularized least-squares problems, where K_p is tied to the norm of the measurement operator $\boldsymbol{\nu}$.

B. Local Lipschitz Continuity for $p \in (1, 2)$

Now, we turn to the case $p \in (1, 2)$ which describes the continuous bridge between sparsity-promoting regularization and Tikhonov regularization. Our mathematical treatment requires one additional assumption

(A4) There is at least one $\tilde{f}$ where the data-fidelity term $E(\mathbf{y}, \boldsymbol{\nu}(\tilde{f}))$ is continuous with respect to its first argument $\mathbf{y}$.

Theorem 4 (Local Lipschitz Continuity of ℓ_p-Regularized Linear Inverse Problems, $p \in (1, 2)$)**.** *Consider the variational problem (18) with $p \in (1, 2)$ and assume that (A1)–(A4) are fulfilled. Then, for any given measurement $\mathbf{y} \in \mathbb{R}^M$, (18) has a unique solution $f_{\mathbf{y}} \in \ell_p$. Moreover, for any compact set $Y \subset \mathbb{R}^M$, there is a constant $r_p(Y) \in \mathbb{R}_+$ such that*

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_p} \leq \frac{(2\, r_p(Y))^{2-p} K_p}{\lambda p(p-1)} \|\mathbf{y}_1 - \mathbf{y}_2\|_2,$$

for any two measurements $\mathbf{y}_1, \mathbf{y}_2 \in Y$.

We thus recover again Lipschitz continuity for $p \to 2$, for which the bound coincides with the one in Theorem 3. For $p \in (1, 2)$, we achieve Lipschitz continuity, albeit only locally in the space $\mathbb{R}^M$ of measurements. Assumption (A4) guarantees the existence of the constant $r_p(Y)$. We defer to Section IV the establishment of an easy-to-evaluate upper bound for this constant.

C. ℓ_p-Regularized Least Squares

Regularized least squares, where the data-fidelity term E is given by (12), is popular in many applications. In this section, we choose ℓ_p-regularized least squares to illustrate our findings expressed in Theorems 3 and 4. For this scenario, (A1) and (A4) hold due to the the properties of the ℓ_2 norm in finite-dimensional spaces. For (A2), we have that E is differentiable with gradient

$$E'(\mathbf{y}) = \boldsymbol{\nabla}_f \{E(\mathbf{y}, \boldsymbol{\nu}(f))\} = \boldsymbol{\nu}^*(\boldsymbol{\nu} f - \mathbf{y}),$$

where $\boldsymbol{\nu}^* \colon \mathbb{R}^M \to \ell_q$ is the adjoint of the measurement operator $\boldsymbol{\nu}$, with $q = (1 - 1/p)^{-1}$. Thus,

$$E'(\mathbf{y}_1) - E'(\mathbf{y}_2) = \boldsymbol{\nu}^*(\mathbf{y}_2 - \mathbf{y}_1)$$

and, therefore,

$$K_p = \|\boldsymbol{\nu}\|_{\mathcal{L}(\ell_p, \mathbb{R}^M)} \coloneqq \sup_{f \in \ell_p} \frac{\|\boldsymbol{\nu}(f)\|_2}{\|f\|_{\ell_p}},$$

where $\|\cdot\|_{\mathcal{L}(\ell_p, \mathbb{R}^M)}$ is the norm of an operator from ℓ_p to $\mathbb{R}^M$. For a generic value of p, this norm might be challenging to compute or approximate. However, if p is between two values for which we know how to compute it, we can use the Riesz-Thorin theorem to bound it from above. For instance, for $p \in (1, 2)$ and $\theta_p = (2 - 2/p)$, we obtain that

$$\|\boldsymbol{\nu}\|_{\mathcal{L}(\ell_p, \mathbb{R}^M)} \leq \|\boldsymbol{\nu}\|^{1-\theta_p}_{\mathcal{L}(\ell_1, \mathbb{R}^M)} \|\boldsymbol{\nu}\|^{\theta_p}_{\mathcal{L}(\ell_2, \mathbb{R}^M)}.$$

1) Revisiting ℓ_2-Regularized Least Squares: For ℓ_2, as for any other Hilbert space $\mathcal{H}$, we have that

$$\|\boldsymbol{\nu}\|_{\mathcal{L}(\ell_2, \mathbb{R}^M)} = \max_{m \in \{1, \dots, M\}} \sqrt{\sigma_m},$$

where σ_m is defined with respect to the Gram matrix $\mathbf{H}$ as in Section II-A. Therefore, for $p = 2$, the bounds in both Theorems 3 and 4 evaluate to

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_2} \leq \max_{m \in \{1, \dots, M\}} \frac{\sqrt{\sigma_m}}{2\lambda} \|\mathbf{y}_1 - \mathbf{y}_2\|_2 . \tag{20}$$

This bound is consistent with (15), but looser. In particular, it does not take into account the assymptotic regime visible in Figure 3, where the norm of the measurement operator dominates. Instead, it characterizes (18) in the strongly regularized regime that corresponds to large λ. We verify this insight providing in Figure 6 a visual comparison of (15) and (20) in terms of λ. This relative behavior is not surprising because Theorems 3 and 4 do not exploit the linearity of the Riesz map, which leads to the closed-form solution (13) and, thereby, to a more nuanced understanding of the effect of the norm of the measurement operator on stability. Here, we note that the behavior observed in Figure 6 is to be expected: increased regularization leads to more stable solution maps. However, the value of the regularization parameter λ still needs to be tuned to obtain the least error, not the most stable reconstruction map.

2) Local Behavior of ℓ_p-Regularized Least Squares for $p \in (1, 2)$: As example of the local behavior ruled by Theorem 4, we now discuss (A4) in more details and exemplify the local constant $r_p(Y)$ when Y is the closed ball $Y = \{\mathbf{y} \in \mathbb{R}^M : \|\mathbf{y}\|_2 \leq \rho\}$. Essentially, $r_p(Y)$ bounds the norm of the optimal solution and (A4) provides a (loose) bound based on the continuity of the overall cost function at $\tilde{f}$ (see Section IV).

Because the least-squares data-fidelity term is continuous for any $f \in \ell_p$, we may choose $\tilde{f} = 0$ for simplicity. This results in $r_p(Y) = (\rho^2/2\lambda)^{1/p}$ and, thus, Theorem 4 implies, for $p \in (1, 2)$ and for $\mathbf{y}_1, \mathbf{y}_2 \in Y$ as above, that

$$\frac{\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\ell_p}}{\|\mathbf{y}_1 - \mathbf{y}_2\|_2} \leq \left(2^{p-1} \frac{\rho^2}{\lambda}\right)^{\frac{2-p}{p}} \frac{1}{\lambda p(p-1)} \|\boldsymbol{\nu}\|_{\mathcal{L}(\ell_p, \mathbb{R}^M)}.$$

IV. Mathematical Formalization

We adopt a functional formulation to prove the results of Theorems 3 and 4 in full generality. The interested reader who would not be acquainted with the terminology is referred to Appendix A, which includes a curated selection of definitions and discussions.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

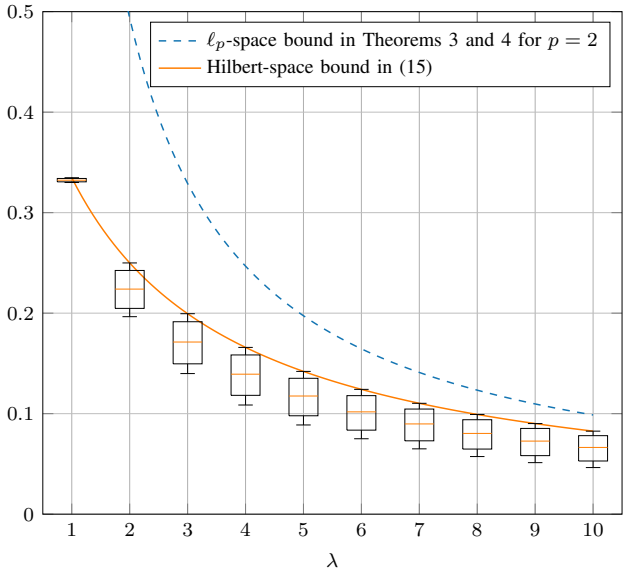

Fig. 6. Lipschitz constant derived in (15) compared to (20), the bound obtained in Theorems 3 and 4 for $p = 2$, in terms of the regularization parameter $\lambda \in \mathbb{R}_+$. The experimental values are as in Figure 3.

We start by formalizing our assumptions and statements with respect to the optimization problem

$$\min_{f \in L_p(\Omega)} \left\{ E\left(\mathbf{y}, \boldsymbol{\nu}(f)\right) + \lambda \|f\|^p_{L_p(\Omega)} \right\}, \tag{21}$$

which generalizes the analysis to function spaces $L_p(\Omega)$ over a domain $\Omega \subset \mathbb{R}^d$. The choice of a countable or finite Ω equipped with the counting measure particularizes (21) to either (2) or (18), respectively.

Similarly to the relationship between (1) and (2), (21) can be seen as the synthesis formulation of the reconstruction problem

$$\min_{\tilde{f} \in L_{p,\mathrm{L}}(\Omega)} \left\{ E\big(\mathbf{y}, \tilde{\boldsymbol{\nu}}(\tilde{f})\big) + \lambda \|\mathrm{L}\tilde{f}\|^p_{L_p(\Omega)} \right\},$$

where the regularization operator L is invertible, $f = \mathrm{L}\tilde{f}$, and $\boldsymbol{\nu} = \tilde{\boldsymbol{\nu}} \circ \mathrm{L}^{-1}$. This type of variational problem has been used, for example, for spline-based interpolation and approximation [47]–[51], inverse diffusion [52], [53], and inverse scattering [54].

In Sections IV-A and IV-B, we present results that are instrumental to complete our proofs.

A. Abstract Characterization of Stability in Variational Inverse Problems

Let us consider the generic optimization problem

$$\min_{f \in \mathcal{X}'} J(\mathbf{y}, f), \tag{22}$$

with a cost functional $J \colon \mathbb{R}^M \times \mathcal{X}' \to \mathbb{R}$. We assume here that the optimization is performed over a Banach space $\mathcal{X}'$ *with a predual* $\mathcal{X}$, in accordance with a more general version of Theorem 1 (see [39]). However, this does not impact the applicability of the results in this section to (21) because $L_p(\Omega)$ spaces are reflexive. For the cost functional J, we assume that

(B1) for any given $\mathbf{y} \in \mathbb{R}^M$, there is a unique solution $f_{\mathbf{y}}$ to (22). Thus, the solution map $S \colon \mathbb{R}^M \to \mathcal{X}'$ with $\mathbf{y} \mapsto f_{\mathbf{y}}$ is well defined.

(B2) The cost function J is differentiable through its second argument, and there is a constant $K > 0$ such that, if

$$e_{i,j} = \nabla_f \left\{ J(\mathbf{y}_i, f) \right\} (f_{\mathbf{y}_j}) \in \mathcal{X}'' \text{ for } i, j \in \{1, 2\}, \tag{23}$$

we have that, for any $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$,

$$\|e_{1,1} - e_{2,1}\|_{\mathcal{X}''} \leq K \|\mathbf{y}_1 - \mathbf{y}_2\|_2 . \tag{24}$$

(B3) There is a subset $Y \subset \mathbb{R}^M$ and constants $\alpha \geq 2$ and $C(Y) \in \mathbb{R}_+$ such that

$$\langle e_{1,1} - e_{1,2}, f_{\mathbf{y}_1} - f_{\mathbf{y}_2} \rangle_{\mathcal{X}'' \times \mathcal{X}'} \geq C(Y) \|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|^\alpha_{\mathcal{X}'} \tag{25}$$

for all $\mathbf{y}_1, \mathbf{y}_2 \in Y$.

Theorem 5 (Local Hölder Stability of (22))**.** *Consider the variational problem (22). Assume (B1)–(B3) are fulfilled. Then,*

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\mathcal{X}'} \leq \left(\frac{K}{C(Y)} \right)^{\frac{1}{\alpha - 1}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2^{\frac{1}{\alpha - 1}} \tag{26}$$

for all $\mathbf{y}_1, \mathbf{y}_2 \in Y$, and S is Hölder continuous on Y.

Proof. From the optimality of $f_{\mathbf{y}_i}$, we have that $e_{i,i} = 0$ for $i \in \{1, 2\}$. Then, we have that

$$\begin{aligned} C(Y) \|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|^\alpha_{\mathcal{X}'} &\leq \langle e_{2,1} - e_{2,2}, f_{\mathbf{y}_1} - f_{\mathbf{y}_2} \rangle_{\mathcal{X}'' \times \mathcal{X}'} && (27) \\ &\leq \|e_{2,1} - e_{2,2}\|_{\mathcal{X}''} \|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\mathcal{X}'} && (28) \\ &= \|e_{2,1} - e_{1,1}\|_{\mathcal{X}''} \|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\mathcal{X}'} && (29) \\ &\leq K \|\mathbf{y}_1 - \mathbf{y}_2\|_2 \|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{\mathcal{X}'}, && (30) \end{aligned}$$

where (27) makes use of (25), (28) follows from the duality bound in Definition 17, (29) is obtained from $e_{1,1} = e_{2,2} = 0$, and (30) makes use of (24). The statement (26) then follows readily by rearrangement of the terms above. □

Remark 6. *The α-uniform convexity of J with respect to f implies (B3). (See [55] for an extensive treatment of uniformly convex functions, and particularly Corollary 3.5.11 for a detailed account of equivalent conditions.)*

B. Growth of the Gradient of L_p Regularizers

In our proofs, we need to verify that (21) fulfills (B3). To do so, we rely on two results on the gradient of the regularizer in (21), namely, on the gradient of the pth power of the L_p norm. In order to simplify the notation, we introduce a useful function in Definition 7.

Definition 7 (Gradient of the L_p Regularizer)**.** *Let $g_p \colon \mathbb{R} \to \mathbb{R}$ be such that*

$$g_p(x) = \operatorname{sign}(x) |x|^{p-1} \text{ for any } x \in \mathbb{R}. \tag{31}$$

Here, $\operatorname{sign}(x) = x/|x|$ for $x \in \mathbb{R} \setminus \{0\}$ and $\operatorname{sign}(0) = 0$.

Then, the gradient of the L_p regularizer evaluates to

$$\nabla_f \left\{ \|f\|^p_{L_p} \right\} = p(g_p \circ f).$$

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

Remark 8. *Since $g_p \circ f$ is proportional to the differential of a functional defined on $L_p(\Omega)$ at $f \in L_p(\Omega)$, we have that $g_p \circ f \in L_q(\Omega)$ for any $f \in L_p(\Omega)$ (see also Remark 23).*

Together, Proposition 9 and Lemma 10 characterize the function in (31) for $p \in (1, \infty)$. First, a known result [56] bounds the growth of g_p for $p \in [2, \infty)$.

Proposition 9 (Growth of g_p for $p \in [2, \infty)$ [56])**.** *The function g_p in Definition 7, for $x, y \in \mathbb{R}$ and $p \in [2, \infty)$, satisfies that*

$$(g_p(x) - g_p(y))(x - y) \geq 2^{2-p}|x - y|^p. \tag{32}$$

Second, we show that g_p also grows controllably for $p \in (1, 2)$.

Lemma 10 (Growth of g_p for $p \in (1, 2)$)**.** *The function g_p in Definition 7, for $x, y \in \mathbb{R}$ and $p \in (1, 2)$, satisfies that*

$$g_p(x) - g_p(y) = (p - 1)\, z^{p-2}\, (x - y) \tag{33}$$

for some $z \leq |x| + |y|$.

Proof. Note that g_p is a differentiable function in $\mathbb{R} \setminus \{0\}$, with derivative $g_p'(x) = (p-1)|x|^{p-2}$.

If $xy \geq 0$ (*i.e.*, x and y have the same sign or one is zero), then (33) corresponds to the statement of the mean-value theorem with $z \in [|x|, |y|]$ and, thus, $z \leq |x| + |y|$.

If $xy < 0$, assume without loss of generality that $x > 0 > y$. Then, by applying the mean-value theorem twice (once between x and 0, and once between 0 and y), we obtain that

$$\Delta g_p = g_p(x) - g_p(y) = (p-1)\big(z_1^{p-2}x - z_2^{p-2}y\big),$$

with $z_1 \in [0, |x|]$ and $z_2 \in [0, |y|]$. Then, let $\alpha = \frac{x}{x-y}$ and note that

$$\Delta g_p = (p-1)\big(\overbrace{z_1^{p-2}\alpha + z_2^{p-2}(1-\alpha)}^{\tilde{\alpha}}\big) \cdot \big(x - y\big).$$

Because $\alpha \in [0, 1]$, we know that $\tilde{\alpha} \in [z_1^{p-2}, z_2^{p-2}]$. Then, by the intermediate-value theorem applied to the continuous function $x \mapsto x^{p-2}$, we know that there is a point $z \in [z_1, z_2]$ such that $z^{p-2} = \tilde{\alpha}$. Furthermore, we have that $z \leq \max_{i \in \{1,2\}} z_i \leq |x| + |y|$. □

C. General Statements and Proofs

We are now equipped to present the generalizations of Theorems 3 and 4 to (21), together with their proofs. In both cases, the proof has the same structure. First, Theorem 1 guarantees (B1) due to the properties of the $L_p(\Omega)$ spaces and assumptions (A1) and (A3). Second, (B2) follows because E has a Lipschitz-continuous gradient (A2) and the regularizer does not depend on $\mathbf{y}$. Finally, we show that (B3) is fulfilled by exploiting the characterization of g_p in Section IV-B. Then, because (B1)–(B3) are fulfilled, we can apply Theorem 5. Note that

$$\nabla_f \{J(\mathbf{y}, f)\} = \nabla_f \{E(\mathbf{y}, \boldsymbol{\nu}(f))\} + \lambda\, p\, g_p(f)\,, \tag{34}$$

where g_p is applied pointwise.

Theorem 11 (Hölder Continuity of L_p-Regularized Linear Inverse Problems, $p \in [2, \infty)$)**.** *Consider the variational problem (21) with $p \in [2, \infty)$ and assume that (A1)–(A3) are fulfilled. Then, for any given measurement $\mathbf{y} \in \mathbb{R}^M$, (21) has a unique solution $f_{\mathbf{y}} \in L_p(\Omega)$. Moreover, it holds that*

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{L_p} \leq \left(\frac{2^{p-2}K_p}{\lambda p}\right)^{\frac{1}{p-1}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2^{\frac{1}{p-1}},$$

for any two $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$.

Proof. Because $\mathcal{X} = L_p(\Omega)$ is a strictly convex space and because $\psi\colon x \mapsto \lambda x^p$ with $\lambda \in \mathbb{R}_+$ is strictly convex, Theorem 1 tells us that (A1) and (A3) imply that (21) has a unique solution $f_{\mathbf{y}}$ and, thus, that (B1) is fulfilled. Assumption (A2) implies that (B2) is fulfilled with constant K_p.

Now, we show that (B3) is fulfilled using Proposition 9. Because E is convex in its second argument, we have that $\langle \nabla_f \{E(\mathbf{y}_1, \boldsymbol{\nu}(f))\}(f_{\mathbf{y}_1}) - \nabla_f \{E(\mathbf{y}_1, \boldsymbol{\nu}(f))\}(f_{\mathbf{y}_2}), f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\rangle_{L_q \times L_p} \geq 0$, and using the decomposition (34) for both $e_{1,1}$ and $e_{1,2}$, we obtain that

$$\begin{aligned}\langle e_{1,1} - e_{1,2}, f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\rangle_{L_q \times L_p} \geq \\ \lambda p \langle g_p(f_{\mathbf{y}_1}) - g_p(f_{\mathbf{y}_2}), f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\rangle_{L_q \times L_p}\,,\end{aligned} \tag{35}$$

where $L_q(\Omega)$ is the Lebesgue function space that identifies with the dual of $L_p(\Omega)$ (see Example 16), with $q = 1/(1 - 1/p)$. Thus, applying Proposition 9 pointwise, we obtain that

$$\langle e_{1,1} - e_{1,2}, f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\rangle_{L_q \times L_p} \geq \lambda p\, 2^{2-p} \|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{L_p}^p$$

and (B3) is satisfied with $\alpha = p$ and $C(Y) = C = \lambda p\, 2^{2-p}$. Thus, Theorem 5 applies and the proof is complete. □

Theorem 12 (Local Lipschitz Continuity of L_p-Regularized Linear Inverse Problems, $p \in (1, 2)$)**.** *Consider the variational problem (21) with $p \in (1, 2)$ and assume that (A1)–(A4) are fulfilled. Then, for any given measurement $\mathbf{y} \in \mathbb{R}^M$, (21) has a unique solution $f_{\mathbf{y}} \in L_p(\Omega)$. Moreover, for any closed and bounded set $Y \subset \mathbb{R}^M$ of measurements, there is a constant $r_p(Y) \in \mathbb{R}_+$ such that*

$$\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{L_p} \leq \frac{(2\, r_p(Y))^{2-p} K_p}{\lambda p(p-1)} \|\mathbf{y}_1 - \mathbf{y}_2\|_2,$$

for any two $\mathbf{y}_1, \mathbf{y}_2 \in Y$.

Proof. By the same arguments as in the proof of Theorem 11, (B1) is fulfilled, (21) has a unique solution $f_{\mathbf{y}}$, and (B2) is fulfilled with constant K_p.

Now, we use (A4) and Lemma 10 to show that (B3) is fulfilled. Consider $\tilde{f} \in L_p(\Omega)$ and recall that (A4) specifies that $E(\cdot, \boldsymbol{\nu}(\tilde{f}))$ is continuous. Then, it holds that

$$\begin{aligned}\lambda \|f_{\mathbf{y}}\|_{L_p(\Omega)}^p &\leq E(\mathbf{y}, \boldsymbol{\nu}(f_{\mathbf{y}})) + \lambda \|f_{\mathbf{y}}\|_{L_p}^p \\ &\leq \max_{\mathbf{y} \in Y} J(\mathbf{y}, \tilde{f}),\end{aligned} \tag{36}$$

for all $\mathbf{y} \in Y$. The last term in (36) is finite because Y is compact and the Weierstrass extreme-value theorem applies. Therefore, there is $r_p(Y) \in \mathbb{R}_+$ such that, for every $\mathbf{y} \in Y$,

$$\|f_{\mathbf{y}}\|_{L_p(\Omega)} \leq r_p(Y) \leq \left(\max_{\mathbf{y} \in Y} J(\mathbf{y}, \tilde{f})/\lambda\right)^{\frac{1}{p}}. \tag{37}$$

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

We now combine (37) with Lemma 10 to show that (B3) is fulfilled. For any two given $\mathbf{y}_1, \mathbf{y}_2 \in Y$, let

- $\xi = |f_{\mathbf{y}_1}| + |f_{\mathbf{y}_2}| \in L_p(\Omega)$;
- $q = 1/(1-1/p)$, the index of the $L_q(\Omega)$ space that identifies with the dual of $L_p(\Omega)$;
- $\beta = p(2-p)/2$;
- $t = 2/(2-p)$ and $s = 2/p = 1/(1-1/t)$.

Then,

$$\begin{aligned}
\|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\|_{L_p}^2 &= \left(\int_\Omega |f_{\mathbf{y}_1} - f_{\mathbf{y}_2}|^p \mathrm{d}\mu\right)^{\frac{2}{p}} \\
&\leq \left(\left\|\xi^\beta\right\|_{L_t} \left\|\frac{|f_{\mathbf{y}_1} - f_{\mathbf{y}_2}|^p}{\xi^\beta}\right\|_{L_s}\right)^{\frac{2}{p}} \qquad (38) \\
&= \|\xi\|_{L_p}^{2-p} \int_\Omega \xi^{p-2} \big(f_{\mathbf{y}_1} - f_{\mathbf{y}_2}\big)^2 \mathrm{d}\mu \\
&\leq \frac{(2\, r_p(Y))^{2-p}}{p-1} \big\langle g(f_{\mathbf{y}_1}) - g(f_{\mathbf{y}_2}), f_{\mathbf{y}_1} - f_{\mathbf{y}_2} \big\rangle_{L_q \times L_p} \qquad (39) \\
&\leq \frac{(2 r_p(Y))^{2-p}}{\lambda p(p-1)} \big\langle e_{1,1} - e_{1,2}, f_{\mathbf{y}_1} - f_{\mathbf{y}_2} \big\rangle_{L_q \times L_p}, \qquad (40)
\end{aligned}$$

for any $e_{1,1}$, $e_{1,2}$. Here, μ represents the Lebesgue measure on $\mathbb{R}^d$. In (38), we use that $1/t + 1/s = 1$ and apply the Hölder inequality for $L_t(\Omega)$ and $L_s(\Omega)$. Then, in (39), we use Lemma 10 pointwise and apply (37). Then, in (40) we conclude, using (35) and the fact that E is convex in its second argument. Thus, (B3) is satisfied with $\alpha = 2$ and $C(Y) = \lambda p(p-1)(2\, r_p(Y))^{p-2}$. Thus, Theorem 5 applies and the proof is complete. □

V. Conclusions

We have shown that ℓ_p-regularized strategies with $p \in (1, \infty)$ in linear inverse problems have good stability properties. The strongest guarantees are those given by Tikhonov regularization in Hilbert spaces (ℓ_2 regularization), for which the reconstruction map is globally Lipschitz continuous. For $p \in (1, 2)$, the reconstruction map is still Lipschitz continuous, albeit only locally in the space of measurements $\mathbb{R}^M$. For $p \in (2, \infty)$, the reconstruction map is globally $1/(p-1)$-Hölder continuous. Thus the stability claim is stronger for p closer to 2. To the best of our knowledge, our bounds are currently the strongest stability results for ℓ_p regularization for $p \in (1, \infty)$. That said, we have not yet investigated the tightness of our bounds for the different regimes, and the option of improved bounds for $p \neq 2$ remains open.

VI. Acknowledgments

The authors would like to thank Dr. Rahul Parhi for fruitful discussions on representer theorems.

We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), École polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE), and Geneva University Hospitals (HUG).

The research leading to these results was partly supported by the European Research Council (ERC) under European Union's Horizon 2020 (H2020), Grant Agreement - Project No 101020573 FunLearn.

Appendix A
Mathematical Preliminaries

Here, we give a digest of the technical definitions and results that are most relevant to our work.

Definition 13 (Banach Space)**.** *A Banach space $\mathcal{X}$ is a vector space with norm $\|\cdot\|_{\mathcal{X}}$ for which it is complete.*

A Banach space is also a complete metric space, as its norm defines a distance $d\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ via $d(f_1, f_2) = \|f_1 - f_2\|_{\mathcal{X}}$ for any two $f_1, f_2 \in \mathcal{X}$. In turn, this distance defines an associated topology—open sets and convergence are defined in terms of the distance d. The completeness must be understood with respect to the distance d, so that Cauchy sequences in the sense of d must converge in $\mathcal{X}$. Banach spaces are instrumental to our work.

Definition 14 (Hilbert Space)**.** *A Hilbert space $\mathcal{H}$ is a Banach space with an inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$. Its norm is induced by the inner product as $\|\varphi\|_{\mathcal{H}} \coloneqq \sqrt{\langle\varphi,\varphi\rangle_{\mathcal{H}}}$ for any $\varphi \in \mathcal{H}$.*

Hilbert spaces are simpler special cases of Banach spaces, in the sense that many of the properties from finite-dimensional Euclidean spaces apply. The inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ allows us to quantify the alignment between two vectors. Specifically, one can define an angle θ between two vectors $\varphi_1, \varphi_2 \in \mathcal{H}$ by

$$\cos(\theta) = \frac{\langle\varphi_1, \varphi_2\rangle_{\mathcal{H}}}{\|\varphi_1\|_{\mathcal{H}} \|\varphi_2\|_{\mathcal{H}}}, \qquad (41)$$

even when these vectors in $\mathcal{H}$ may be infinite-dimensional.

Definition 15 (Continuous Dual of a Banach Space)**.** *The dual space $\mathcal{X}'$ of a Banach space $\mathcal{X}$ is the vector space formed by all linear and continuous functionals $\nu\colon \mathcal{X} \to \mathbb{R}$. The corresponding values $\nu(f)$ are often expressed in terms of the bilinear form (the duality product) $\langle\cdot,\cdot\rangle_{\mathcal{X}'\times\mathcal{X}}\colon \mathcal{X}' \times \mathcal{X} \to \mathbb{R}$ given by $\langle\nu, f\rangle_{\mathcal{X}'\times\mathcal{X}} \coloneqq \nu(f)$. The dual space $\mathcal{X}'$ is, in turn, a Banach space, when equipped with the operator norm*

$$\|\nu\|_{\mathcal{X}'} \coloneqq \sup_{f \in \mathcal{X}\setminus\{0\}} \frac{|\langle\nu, f\rangle_{\mathcal{X}'\times\mathcal{X}}|}{\|f\|_{\mathcal{X}}}. \qquad (42)$$

Example 16 ($L_q(\Omega)$ is the Dual of $L_p(\Omega)$ with $q = p/(p-1)$, $p, q \in (1, \infty)$)**.** *The concept of a dual vector space can be made more concrete by identifying it with another vector space. Most relevant to our work is $\mathcal{X} = L_p(\Omega)$. Then, the Riesz theorem asserts that every $\nu \in \mathcal{X}'$ can be expressed, $\forall f \in \mathcal{X}$, as*

$$\langle\nu, f\rangle_{\mathcal{X}'\times\mathcal{X}} = \int_\Omega \tilde{\nu}\, f \,\mathrm{d}\mu$$

for some $\tilde{\nu} \in L_q(\Omega)$, and $\|\nu\|_{\mathcal{X}'} = \|\tilde{\nu}\|_{L_q(\Omega)}$. Because of this one-to-one isomorphism, we abuse the notation and write that $\mathcal{X}' = L_q(\Omega)$ and $\tilde{\nu} = \nu$. This is common practice but it is to handle with care, as these are different mathematical objects.

Proposition 17 (Duality Bound)**.** *For any $\nu \in \mathcal{X}'$ and $f \in \mathcal{X}$, one has that*

$$|\langle\nu, f\rangle_{\mathcal{X}'\times\mathcal{X}}| \leq \|\nu\|_{\mathcal{X}'} \|f\|_{\mathcal{X}},$$

which is called the duality bound. This bound is sharp for any dual pair of Banach spaces $(\mathcal{X}, \mathcal{X}')$.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

Proof. The duality bound follows immediately from (42). Sharpness is a corollary of the Hahn-Banach theorem. □

The duality bound generalizes the Cauchy-Schwartz inequality in Hilbert spaces

$$|\langle \varphi_1, \varphi_2 \rangle_{\mathcal{H}}| \leq \|\varphi_1\|_{\mathcal{H}} \|\varphi_2\|_{\mathcal{H}}, \forall \varphi_1, \varphi_2 \in \mathcal{H}.$$

The Cauchy-Schwartz inequality is saturated by parallel vectors, corresponding to $|\cos(\theta)| = 1$ in (41). Analogously, it is worthwhile to identify the set of dual vectors that saturate the duality bound for a given vector $\nu \in \mathcal{X}$.

Definition 18 (Duality Map)**.** *The duality map is the set-valued map $\mathcal{J}_{\mathcal{X}} \colon \mathcal{X} \rightrightarrows \mathcal{X}'$ given by*

$$\mathcal{J}_{\mathcal{X}}(f) = \left\{ \nu \in \mathcal{X}' : \begin{array}{l} \|\nu\|_{\mathcal{X}'} = \|f\|_{\mathcal{X}} \text{ and} \\ \langle \nu, f \rangle_{\mathcal{X}' \times \mathcal{X}} = \|\nu\|_{\mathcal{X}'} \|f\|_{\mathcal{X}} \end{array} \right\}.$$

As we have seen in Theorem 1, the duality map characterizes the set of solutions of the variational problems in Banach spaces that take the form (5). For strictly convex Banach spaces $\mathcal{X}$, the duality map is single-valued.

Proposition 19 (Duality Map in $L_q(\Omega)$, $q \in (1, \infty)$ [57, Corollary 4.10])**.** *The duality map for $\mathcal{X}' = L_q(\Omega)$ with $q \in (1, \infty)$ is single-valued and given by*

$$\mathcal{J}_{\mathcal{X}'}(\nu) = \frac{|\nu|^{q-1}}{\|\nu\|_{L_q}^{q-2}} \operatorname{sign}(\nu), \tag{43}$$

where the absolute value $|\cdot|$ and $\operatorname{sign}(\cdot)$ operators are applied element-wise.

The finite-dimensional discrete equivalent of (43) was implemented to obtain the visualizations in Figures 1 and 2. There, we see that the effect of the duality map for $\mathcal{X}' = \ell_q$ resembles the behavior that we expect from ℓ_p-norm regularization in variational problems for $1/p + 1/q = 1$.

Definition 20 (Riesz Map and Hilbert Spaces)**.** *For a Hilbert space $\mathcal{X} = \mathcal{H}$, the duality map $\mathcal{J}_{\mathcal{X}}$ is single-valued. Furthermore, its inverse $\mathcal{R} = \mathcal{J}_{\mathcal{X}}^{-1} \colon \mathcal{H}' \to \mathcal{H}$ is called the Riesz map. It maps a dual vector $\nu \in \mathcal{H}'$ to its Riesz representer $\tilde{\nu} \in \mathcal{H}$, such that*

$$\langle \nu, f \rangle_{\mathcal{H}' \times \mathcal{H}} = \langle \tilde{\nu}, f \rangle_{\mathcal{H}}$$

for all $f \in \mathcal{H}$.

This is the equivalent of the phenomenon described in Example 16, but in Hilbert spaces.

Definition 21 (Reflexive Banach Space)**.** *A Banach space $\mathcal{X}$ is reflexive if it can be identified with its bidual $\mathcal{X}''$.*

The bidual space $\mathcal{X}''$ is simply the continuous dual of $\mathcal{X}'$: that is, it is formed of all the continuous linear functionals $\tilde{f} \colon \mathcal{X}' \to \mathbb{R}$. In particular, we can construct such a functional from any $f \in \mathcal{X}$ through the identity

$$\langle \tilde{f}, \nu \rangle_{\mathcal{X}'' \times \mathcal{X}'} = \langle \nu, f \rangle_{\mathcal{X}' \times \mathcal{X}},$$

which specifies the canonical embedding of $\mathcal{X}$ into $\mathcal{X}''$. Although not formally accurate in the same sense as Example 16, this is usually denoted as $\mathcal{X} \subseteq \mathcal{X}''$. In reflexive Banach spaces, it is in the same sense that $\mathcal{X} = \mathcal{X}''$: all linear and continuous functionals on $\mathcal{X}'$ can be represented by an element of $\mathcal{X}$ through the canonical embedding, which is then one-to-one.

For a general Banach space $\mathcal{X}$, a more general version of Theorem 1 (*cf.* [39]) restricts the choice of the measurement functionals (of $f \in \mathcal{X}'$) by assuming that $\{\nu_m\}_{m=1}^M \subset \mathcal{X} \subseteq \mathcal{X}''$. However, this is not a limitation in reflexive Banach spaces because $\mathcal{X} = \mathcal{X}''$, which yields the formulation of Theorem 1 in this paper. In that more general version of Theorem 1, E and ψ are also not required to be strictly convex. Then, the solution is no longer unique. Instead, the solution set is guaranteed to be nonempty, convex, and weak*-compact. We include here the definition of weak*-compact for completeness.

Definition 22 (Weak* Compactness)**.** *A weak*-compact set in the dual of a separable Banach space $\mathcal{X}$ is a set $\mathcal{C} \subset \mathcal{X}'$ such that, for any sequence $\{\nu_n\}_{n=1}^{\infty} \subset \mathcal{C}$, there is a weak*-convergent subsequence $\{\nu_{n_r}\}_{r=1}^{\infty}$, meaning that*

$$\langle \nu_{n_r}, f \rangle_{\mathcal{X}' \times \mathcal{X}} \to \langle \nu, f \rangle_{\mathcal{X}' \times \mathcal{X}} \text{ as } r \to \infty$$

for some $\nu \in \mathcal{C}$ and any $f \in \mathcal{X}$.

Weak* compactness is useful in variational theory because it guarantees the existence of minimizers of certain cost functionals using the generalized Weierstrass extreme-value theorem. This makes the result in [39] more attractive, as the further selection among the solutions of an initial variational problem is made possible through variational techniques.

Remark 23 (Gradient of a Functional)**.** *The gradient of a differentiable functional (in the Fréchet sense) $J \colon \mathcal{X}' \to \mathbb{R}$ is the map $\nabla J \colon \mathcal{X}' \to \mathcal{X}''$ such that*

$$\lim_{\|\tilde{f}\|_{\mathcal{X}'} \to 0} \left\{ \frac{J(f + \tilde{f}) - J(f) - \langle \nabla J(f), \tilde{f} \rangle_{\mathcal{X}'' \times \mathcal{X}'}}{\|\tilde{f}\|_{\mathcal{X}'}} \right\} = 0.$$

This contextualizes the proof of Theorem 5, where we explicitly treat gradient and subdifferential values as elements of $\mathcal{X}''$. In the main body of the paper, we can avoid this explicit treatment because of the reflexivity of the spaces being considered.

References

[1] K. Jin, M. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[2] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[3] R. Strack, "AI transforms image reconstruction," *Nature Methods*, vol. 15, no. 5, pp. 309–309, 2018.

[4] P. del Aguila Pla, V. Saxena, and J. Jaldén, "Spotnet – Learned iterations for cell detection in image-based immunoassays," in *Proceedings of the Sixteenth IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'19)*, Venice, Italian Republic, Apr. 2019, pp. 1023–1027.

[5] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 088–30 095, 2020.

[6] P. Hagemann and S. Neumayer, "Stabilizing invertible neural networks using mixture models," *Inverse Problems*, vol. 37, no. 8, p. 085002, 2021.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2023.3236161

[7] N. M. Gottschling, V. Antun, B. Adcock, and A. C. Hansen, "The troublesome kernel: Why deep learning for inverse problems is typically unstable," Jan. 2020, archived as arXiv:2001.01258.

[8] M. Z. Darestani, A. S. Chaudhari, and R. Heckel, "Measuring robustness in deep learning based compressive sensing," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 2433–2444.

[9] M. Genzel, J. Macdonald, and M. Marz, "Solving inverse problems with deep neural networks—Robustness included," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[10] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, ser. Mathematics and Its Applications. Kluwer Academic Publishers, 1996, no. 375.

[11] C. R. Vogel, *Computational methods for inverse problems*, ser. Frontiers in Applied Mathematics. SIAM, Philadelphia, 2002, vol. 23.

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

[13] T. Schuster, B. Katenbacher, B. Hofmann, and K. S. Kazimierski, *Regularization methods in Banach spaces*, ser. Radon Series on Computational and Applied Mathematics. De Gruyter, 2012, no. 10.

[14] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, pp. 1–111, 2018.

[15] M. T. McCann and M. Unser, "Biomedical image reconstruction: From the foundations to deep neural networks," *Foundations and Trends® in Signal Processing*, vol. 13, no. 3, pp. 283–359, 2019.

[16] A. N. Tikhonov, "On the stability of inverse problems," *Proceedings of the USSR Academy of Sciences*, vol. 39, pp. 176–179, 1943.

[17] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, ser. Scripta Series in Mathematics. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York-Toronto, 1977, translated from Russian.

[18] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[19] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[20] C. Chen and O. Öktem, "Indirect image registration with large diffeomorphic deformations," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 575–617, 2018.

[21] S. Neumayer, J. Persch, and G. Steidl, "Regularization of inverse problems via time discrete geodesics in image spaces," *Inverse Problems*, vol. 35, no. 5, p. 055005, 2019.

[22] L. Lang, S. Neumayer, O. Öktem, and C.-B. Schönlieb, "Template-based image reconstruction from sparse tomographic data," *Applied Mathematics and Optimization*, vol. 82, pp. 1081–1109, 2019.

[23] S. Neumayer and A. Topalovic, "Template based image reconstruction facing different topologies," Aug. 2021, Archived as arXiv:2108.13777.

[24] W. J. Fu, "Penalized regressions: The bridge versus the LASSO," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, pp. 397–416, 1998.

[25] S. Wang, H. Weng, and A. Maleki, "Which bridge estimator is the best for variable selection?" *The Annals of Statistics*, vol. 48, no. 5, pp. 2791–2823, 2020.

[26] M. Kloft and G. Blanchard, "On the convergence rate of ℓ_p-norm multiple kernel learning," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2465–2502, 2012.

[27] M. Burger and S. Osher, "Convergence rates of convex variational regularization," *Inverse Problems*, vol. 20, no. 5, pp. 1411–1421, 2004.

[28] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[29] E. Resmerita and O. Scherzer, "Error estimates for non-quadratic regularization and the relation to enhancement," *Inverse Problems*, vol. 22, no. 3, pp. 801–814, 2006.

[30] M. Grasmair, M. Haltmeier, and O. Scherzer, "Sparse regularization with ℓ^q penalty term," *Inverse Problems*, vol. 24, no. 5, pp. 055 020, 13, 2008.

[31] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, ser. Springer Series in Operations Research. Springer, New York, 2000.

[32] S. Durand and M. Nikolova, "Stability of the minimizers of least squares with a non-convex regularization. Part I: Local behavior," *Applied Mathematics and Optimization*, vol. 53, no. 2, pp. 185–208, 2006.

[33] I. Shvartsman, "On stability of minimizers in convex programming," *Nonlinear Analysis*, vol. 75, no. 3, pp. 1563–1571, 2012.

[34] H. Xu, C. Caramanis, and S. Mannor, "Sparse algorithms are not stable: A no-free-lunch theorem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 187–193, 2012.

[35] M. Unser, J. Fageot, and H. Gupta, "Representer theorems for sparsity-promoting ℓ_1 regularization," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5167–5180, 2016.

[36] H. Gupta, J. Fageot, and M. Unser, "Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization," *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4670–4684, 2018.

[37] M. Unser, "A representer theorem for deep neural networks," *Journal of Machine Learning Research*, vol. 20, no. 110, pp. 1–30, 2019.

[38] M. Unser and S. Aziznejad, "Convex optimization in sums of Banach spaces," *Applied and Computational Harmonic Analysis*, vol. 56, pp. 1–25, 2022.

[39] M. Unser, "A unifying representer theorem for inverse problems and machine learning," *Foundations of Computational Mathematics*, vol. 21, no. 4, pp. 941–960, 2021.

[40] H. Zhang and J. Zhang, "Regularized learning in Banach spaces as an optimization problem: Representer theorems," *Journal of Global Optimization*, vol. 54, no. 2, pp. 235–250, 2012.

[41] A. Flinth and P. Weiss, "Exact solutions of infinite dimensional total-variation regularized problems," *Information and Inference: A Journal of the IMA*, vol. 8, no. 3, pp. 407–443, 2018.

[42] C. Boyer, A. Chambolle, Y. D. Castro, V. Duval, F. de Gournay, and P. Weiss, "On representer theorems and convex regularization," *SIAM Journal on Optimization*, vol. 29, no. 2, pp. 1260–1281, 2019.

[43] K. Bredies and M. Carioni, "Sparsity of solutions for variational inverse problems with finite-dimensional data," *Calculus of Variations and Partial Differential Equations*, vol. 59, no. 1, p. 14, 2020.

[44] F. Bartolucci, E. De Vito, L. Rosasco, and S. Vigogna, "Understanding neural networks with reproducing kernel Banach spaces," Sep. 2021, archived as arXiv:2109.09710v1.

[45] M. Grasmair, M. Haltmeier, and O. Scherzer, "Necessary and sufficient conditions for linear convergence of ℓ^1-regularization," *Communications on Pure and Applied Mathematics*, vol. 64, no. 2, pp. 161–182, 2011.

[46] J. Flemming, B. Hofmann, and I. Veselić, "A unified approach to convergence rates for ℓ^1-regularization and lacking sparsity," *Journal of Inverse and Ill-Posed Problems*, vol. 24, no. 2, pp. 139–148, 2016.

[47] C. de Boor, "On "best" interpolation," *Journal of Approximation Theory*, vol. 16, no. 1, pp. 28–42, 1976.

[48] C. Micchelli, P. W. Smith, J. Swetits, and J. D. Ward, "Constrained L_p approximation," *Constructive Approximation*, vol. 1, no. 1, pp. 93–102, 1985.

[49] J. E. Lavery, "Univariate cubic L_p splines and shape-preserving, multiscale interpolation by univariate cubic L_1 splines," *Computer Aided Geometric Design*, vol. 17, no. 4, pp. 319–336, 2000.

[50] P. Auquiert, O. Gibaru, and E. Nyiri, "C^1 and C^2-continuous polynomial parametric L_p splines ($p \geq 1$)," *Computer Aided Geometric Design*, vol. 24, no. 7, pp. 373–394, 2007.

[51] P. Bohra and M. Unser, "Continuous-domain signal reconstruction using L_p-norm regularization," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4543–4554, 2020.

[52] P. del Aguila Pla and J. Jaldén, "Cell detection by functional inverse diffusion and non-negative group sparsity—Part I: Modeling and Inverse problems," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5407–5421, 2018.

[53] ——, "Cell detection by functional inverse diffusion and non-negative group sparsity—Part II: Proximal optimization and Performance evaluation," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5422–5437, 2018.

[54] A. Lechleiter, K. S. Kazimierski, and M. Karamehmedović, "Tikhonov regularization in L^p applied to inverse medium scattering," *Inverse Problems*, vol. 29, no. 7, p. 075003, 2013.

[55] C. Zălinescu, *Convex analysis in general vector spaces*. World Scientific Publishing Co., River Edge, 2002.

[56] ——, "On uniformly convex functions," *Journal of Mathematical Analysis and Applications*, vol. 95, no. 2, pp. 344–374, 1983.

[57] I. Cioranescu, *Geometry of Banach spaces, duality mappings and nonlinear problems*, ser. Mathematics and Its Applications. Kluwer Academic Publishers, 1990, no. 62.

[58] F. Natterer, *The mathematics of computerized tomography*. Wiley, 1986.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on January 18,2023 at 09:37:14 UTC from IEEE Xplore. Restrictions apply.