

# **A General Framework for Soft-Shrinkage with Applications to Blind Deconvolution and Wavelet Denoising**

**R. Ramlau, L. Justen**

**RICAM-Report 2007-19**

# A General Framework for Soft-Shrinkage with Applications to Blind Deconvolution and Wavelet Denoising

L Justen<sup>1</sup> and R Ramlau<sup>2</sup>

September 14, 2007

## Abstract

We consider the abstract problem of approximating a function  $\psi^0 \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  given only noisy data  $\psi^\delta \in \mathbb{L}^2(\mathbb{R}^d)$ . We recall that minimization of the corresponding Tikhonov functional leads to continuous soft-shrinkage and prove convergence results. If the noise-free data  $\psi^0$  belongs to the source space  $\mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  for some  $0 < u < 1$ , we show convergence rates, which are order-optimal. We consider a-priori parameter choice rules as well as the discrepancy principle, which is shown to be order-optimal as well. We then introduce a framework by combining soft-shrinkage with a linear invertible isometry and show that the results obtained for the abstract minimization problem can be transferred to applications such as blind deconvolution and wavelet denoising.

## 1 Introduction

In many practical applications one has to extract information out of measured data. Due to the measurement process, we have no access to the exact data  $\psi^0$  but rather to its noisy version  $\psi^\delta$ . Additionally, the noisy data usually belongs to a larger function space than the exact data, which makes it sometimes difficult to extract the information searched for. E.g., an image with piecewise continuous gray value distribution will belong to the Sobolev spaces  $H^s$  with  $s < 1/2$ , whereas its noisy version will usually belong to  $\mathbb{L}^2 = H^0$  only, which causes problems if one wants to extract edges. A common way to overcome these difficulties is to remove the noise by some data denoising techniques and / or to compute an approximation of the exact data that belongs to the proper (smaller) function space. Denoising techniques are also useful for the inversion of ill - posed problems of type  $y = F(x)$ : Although the operator  $F$  might be continuously invertible on  $R(F)$ , it is usually not continuously invertible on the space that contains the noisy data. This is e.g. the case for the Radon transform [1, 2] or the blind deconvolution problem, which will be considered in Section 8. A way to obtain a stable approximation to the solution of the equation is by first smoothing the data into the range of the operator and then to apply the inverse operator. These two step methods (data denoising + inversion) were investigated in [3, 2, 4, 5, 6].

In this paper, we are in particular interested in exact data  $\psi^0$  that belongs to the function space  $\mathbb{L}^1(\mathbb{R})$  and where the noisy data  $\psi^\delta$  belongs only to the space  $\mathbb{L}^2(\mathbb{R})$ . The computation of an  $\mathbb{L}^1$ -approximation  $\psi_\alpha^\delta$  to the true data is obtained by minimizing the Tikhonov functional

$$F_\alpha(\psi) = \frac{1}{2} \|\psi^0 - \psi^\delta\|_2^2 + \alpha \|\psi\|_1 .$$

It turns out that the minimizer of  $F_\alpha$  can be computed by a point-wise shrinkage operation. The main part of the paper will be concerned with the investigation of the regularization properties of the minimizer of the functional. We will propose an a priori parameter rule as well as an a posteriori parameter choice rule, namely Morozov's discrepancy principle, and will show that both rules produce a convergent scheme, i.e.  $\psi_\alpha^\delta \rightarrow \psi$  as  $\delta \rightarrow 0$ . Provided that the true data  $\psi^0$  belongs to the space  $\mathbb{L}^{1-u}$  for  $0 < u < 1$ , we will prove convergence rates for both parameter choice rules that are also of optimal order. By using a suitable transformation, these results will be extended to other important data denoising problems, e.g. continuous Fourier shrinkage, Besov or sparsity type penalties and wavelet shrinkage. To our knowledge, this is the first time that convergence and convergence rates for an a posteriori parameter rule are available for Tikhonov regularization with a non - quadratic penalty term, even in the case that only the identity is used as operator.

As data denoising/estimation has been of interest for some time, there exists an abundant literature on the topic. We will focus our review to variational methods and wavelet based approaches. The variational approaches are based on the Tikhonov type functional

$$J_\alpha(\psi) = \frac{1}{2} \|\psi^0 - \psi^\delta\|_2^2 + \alpha \Omega(\psi) ,$$

where  $\Omega$  is a suitable positive and convex functional. The simplest choice for the functional is  $\Omega(\psi) = \|\psi\|_X^2$ , where  $X$  is a Hilbert space. The analysis of this functional reduces to standard Tikhonov regularization with the identity as operator, and therefore also convergence and convergence rates of the method are available, see e.g. [2, 7]. Although the method is numerically easy to handle, its main drawback is its frequent oversmoothing of the data. Another very popular choice for the penalty, in particular to preserve edges, is the total variation seminorm

$$\Omega(\psi) = \int |\nabla \psi(x)| dx .$$

It has been introduced in [8] for the reconstruction of images with edges. The problem formulation given in [8] already includes a discrepancy principle. Within the last decade, many contributions have been made to TV related image reconstruction, e.g. [9, 10]. However, only little is known on convergence and convergence rates for the method: In [11] a convergence rate result with respect to the  $L^2$  norm is derived for an a priori parameter rule, and [12] gives a rate result in terms of a Bregman distance. Another drawback of the method are the numerical difficulties of the minimization of the functional, a problem that was addressed e.g. in [13, 14].

The choice of a Besov norm penalty,  $\Omega(\psi) = \|\psi\|_{B_{pq}^s}^{(q)}$  allows a better fine-tuning of the reconstruction to the smoothness properties of the signal. The functional has been extensively studied in [15, 16, 17]. In contrast to TV regularization, the minimizer of the functional with Besov penalty is usually easy to compute (at least for the case  $p = q$ ). Moreover, a proper chosen Besov penalty (e.g. the  $B_{11}^1$  norm in two dimensions) will produce TV like reconstructions, see e.g. [18, 19, 20]. As we will show in Section 4, our results concerning the use of Morozov's discrepancy principle and the associated convergence rates will carry over to certain types of Besov penalties, which will make this rule accessible for Besov type denoising.

A standard denoising procedure is denoising via wavelet shrinkage. In this approach, a signal is decomposed in a wavelet basis, and small coefficients - assuming they carry the noise - are set to zero, whereas the other coefficients remain unaltered (hard shrinkage) or will be damped (soft shrinkage). A first analysis of the method was done in [21, 22, 23]. More recently, wavelet shrinkage has also been interpreted as the result of various descent problems [24]. To our knowledge, no convergence rate results for a posteriori parameter choice rules comparable to the discrepancy principle are known for wavelet shrinkage. Additionally, many of the available error estimates are given in a stochastic setting, e.g. [25].

The paper is organized as follows: In Section 2 we discuss the minimization of the Tikhonov functional with  $\mathbb{L}^1$ -penalty, introduce a priori / a posteriori parameter choice rules and give convergence and convergence rate results. Section 3 discusses the order optimality of the presented convergence rates, whereas Section 4 extends the obtained results for the  $\mathbb{L}^1$ -penalty to functionals that are obtained via suitable transformations. In Section 5 these results will be applied to continuous Fourier shrinkage, and in Section 6 to wavelet shrinkage. Finally, Sections 7 and 8 present numerical results for the blind deconvolution problem and wavelet denoising.

## 2 Continuous soft-shrinkage

Let  $\psi^0 \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  denote an arbitrary function that denotes the noise-free data. Assume we are given noisy data  $\psi^\delta$  with  $\|\psi^\delta - \psi^0\|_2 \leq \delta$  and we would like to find regularized data  $\psi_\alpha^\delta$  such that the error  $\|\psi_\alpha^\delta - \psi^0\|_1$  is small. A standard way is to consider the Tikhonov functional

$$F_\alpha(\psi) = \frac{1}{2}\|\psi - \psi^\delta\|_2^2 + \alpha\|\psi\|_1 \quad (1)$$

and to define the regularized solution by

$$\psi_\alpha^\delta := \underset{\psi}{\operatorname{argmin}} F_\alpha(\psi).$$

Note that throughout this paper,  $\|\cdot\|_p$  denotes the norm on the whole space  $\mathbb{L}^p(\mathbb{R}^d)$ . Norms on  $\mathbb{L}^p(\Omega)$  for some domain  $\Omega \subset \mathbb{R}^d$  are denoted by  $\|\cdot\|_{\mathbb{L}^p(\Omega)}$ .

We will proceed as follows: First, we will show that  $\psi_\alpha^\delta$  can be explicitly computed by *soft-shrinkage* of the data  $\psi^\delta$ . Then, we will show that under

an appropriate a-priori parameter choice  $\alpha = \alpha(\delta)$  the regularized solution  $\psi_\alpha^\delta$  converges to  $\psi^0$  in  $\mathbb{L}^1(\mathbb{R}^d)$  for  $\delta \rightarrow 0$ . Moreover, given a certain source representation of the true solution  $\psi^0$ , we will show convergence rates. Finally, we will analyze an a-posteriori parameter choice rule, namely the discrepancy principle.

We extend the functional  $F_\alpha$  to  $\mathbb{L}^2(\mathbb{R}^d)$  by setting  $F_\alpha(\psi) = \infty$  whenever  $\psi \in \mathbb{L}^2(\mathbb{R}^d) \setminus \mathbb{L}^1(\mathbb{R}^d)$ . Then,  $F_\alpha$  is proper, convex and lower semicontinuous on  $\mathbb{L}^2(\mathbb{R}^d)$ , see [24] and [26]. Thus,  $\psi_\alpha^\delta$  is a minimizer of  $F_\alpha$  if and only if

$$0 \in \partial F_\alpha(\psi_\alpha^\delta), \quad (2)$$

where  $\partial F_\alpha(\psi)$  is the subgradient of  $F_\alpha$  at  $\psi$ . The subgradient is given by

$$\partial F_\alpha(\psi) = \psi - \psi^\delta + \alpha \text{Sign}(\psi). \quad (3)$$

The set-valued function  $\text{Sign}(\psi)$  is the subgradient of  $\partial \|\cdot\|_1$  at  $\psi$ . It contains all functions that belong to  $\text{sign}(\psi(\omega))$  (see below) in a point-wise manner [24], i.e.

$$\text{Sign}(\psi) = \{\phi \in \mathbb{L}^2(\mathbb{R}^d) \mid \phi(\omega) \in \text{sign}(\psi(\omega)) \text{ for a.e. } \omega \in \mathbb{R}^d\}, \quad (4)$$

where the set-valued signum-function  $\text{sign} \psi(\omega)$  is the subgradient of the function  $z \mapsto |z|$  at  $z = \psi(\omega)$ . It is given by

$$\text{sign}(z) = \begin{cases} \left\{ \frac{z}{|z|} \right\} & \text{if } z \neq 0, \\ \{\xi \in \mathbb{C} \mid |\xi| \leq 1\} & \text{otherwise.} \end{cases}$$

The subgradient  $\text{Sign}(\psi) = \partial \|\psi\|_1$  for some  $\psi$  is illustrated in Figure 1. For simplicity of notation we will identify sets with a single element  $\left\{ \frac{z}{|z|} \right\}$  with the element itself. Equations (2)-(4) now imply

$$\psi_\alpha^\delta(\omega) \in \psi^\delta(\omega) - \alpha \text{sign} \psi_\alpha^\delta(\omega) \quad \text{for a.e. } \omega \in \mathbb{R}^d,$$

and thus

$$\psi_\alpha^\delta(\omega) = \max\{|\psi^\delta(\omega)| - \alpha, 0\} \text{sign} \psi^\delta(\omega) =: (S_\alpha \psi^\delta)(\omega). \quad (5)$$

That is, the minimizer  $\psi_\alpha^\delta$  of the Tikhonov functional (1) is simply a shrunk version of  $\psi^\delta$ . The operator  $S_\alpha$  is called (continuous) *soft-shrinkage operator*. Minimization of  $F_\alpha$  corresponds to continuous soft-shrinkage of  $\psi^\delta$ .

## 2.1 A-priori parameter choice rules

We show that the regularized solution  $\psi_\alpha^\delta = S_\alpha \psi^\delta$  converges to the true solution  $\psi^0$  in  $\mathbb{L}^1(\mathbb{R}^d)$  for  $\delta \rightarrow 0$  given an appropriate a-priori parameter choice rule  $\alpha = \alpha(\delta)$ . We will need that  $S_\alpha$  and  $I - S_\alpha$  are non-expansive operators:

**Lemma 2.1.** *For any  $1 \leq p \leq \infty$ , the operators  $S_\alpha$  and  $I - S_\alpha$  are non-expansive on  $\mathbb{L}^p$ , i.e. for every  $\psi_1, \psi_2 \in \mathbb{L}^p(\mathbb{R}^d)$  the inequalities*

$$\|S_\alpha \psi_1 - S_\alpha \psi_2\|_p \leq \|\psi_1 - \psi_2\|_p \quad \text{and} \quad (6)$$

$$\|\psi_1 - S_\alpha \psi_1 - (\psi_2 - S_\alpha \psi_2)\|_p \leq \|\psi_1 - \psi_2\|_p \quad (7)$$

hold true.

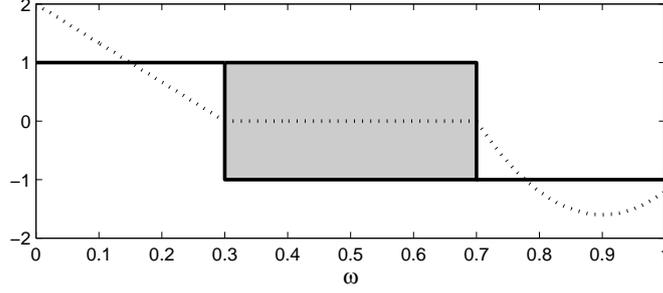


Figure 1: Some function  $\psi$  (dotted) and the real part of the subgradient  $\text{Sign}(\psi) = \partial\|\psi\|_1$  (solid black and gray). Note that  $\psi(\omega) = 0$  for  $\omega \in [0.3, 0.7]$ , so that the set-valued subgradient takes all values between  $-1$  and  $1$  in that interval.

*Proof.* The proof is elementary and can be found in [27].  $\square$

For  $0 < u \leq 1$ , the condition  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  will imply a certain convergence rate. The following lemma collects some facts needed to prove the convergence results. It will turn out that we need the sets

$$\Omega_\alpha := \{\omega \in \mathbb{R}^d \mid |\psi^0(\omega)| > \alpha\} \quad \text{and} \quad \Omega_\alpha^\delta := \{\omega \in \mathbb{R}^d \mid |\psi^\delta(\omega)| > \alpha\}.$$

Using the notation

$$|\Omega| = \int_{\Omega} d\omega$$

we obtain the following result.

**Lemma 2.2.** *Given  $0 < t \leq 1$ , let  $\psi^0 \in \mathbb{L}^t(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $\psi^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  with  $\|\psi^0 - \psi^\delta\|_2 \leq \delta$ . Then, the following holds:*

- (a).  $\|\psi^0\|_t^t = \int_0^\infty t\alpha^{t-1}|\Omega_\alpha| d\alpha$ .
- (b).  $|\Omega_\alpha| = o(\alpha^{-t})$  as  $\alpha \rightarrow 0$ .
- (c).  $\int_{\mathbb{R}^d \setminus \Omega_\alpha} |\psi^0(\omega)| d\omega \leq \frac{1}{t}\alpha^{1-t}\|\psi^0\|_t^t$ .
- (d).  $|\Omega_\alpha^\delta| \leq |\Omega_{\alpha/2}| + 4\frac{\delta^2}{\alpha^2}$ .
- (e).  $\mathbb{R}^d \setminus \Omega_\alpha^\delta \subset (\Omega_{2\alpha} \setminus \Omega_\alpha^\delta) \cup (\mathbb{R}^d \setminus \Omega_{2\alpha})$ .
- (f).  $|\Omega_{2\alpha} \setminus \Omega_\alpha^\delta| \leq \frac{\delta^2}{\alpha^2}$ .

*Proof.* (a) By Fubini's Theorem, we have

$$\|\psi^0\|_t^t = \int_{\mathbb{R}^d} |\psi^0(\omega)|^t d\omega = \int_{\mathbb{R}^d} \int_0^{|\psi^0(\omega)|} t\alpha^{t-1} d\alpha d\omega = \int_0^\infty t\alpha^{t-1}|\Omega_\alpha| d\alpha.$$

- (b) Assume that the statement is not true. Then, there exists a constant  $C > 0$  and a sequence  $0 < \alpha_n \rightarrow 0$ , such that

$$|\Omega_{\alpha_n}| \geq C\alpha_n^{-t}$$

for all  $n$ . By choosing a suitable subsequence if necessary we may assume that  $\alpha_{n+1} \leq \alpha_n/2$ . Noting that  $|\Omega_\alpha|$  decreases as  $\alpha$  increases, we obtain the contradiction

$$\begin{aligned} \infty > \|\psi^0\|_t^t &\stackrel{(a)}{=} \int_0^\infty t\alpha^{t-1}|\Omega_\alpha| d\alpha \geq \sum_{n=1}^\infty \int_{\alpha_{n+1}}^{\alpha_n} t\alpha^{t-1}|\Omega_\alpha| d\alpha \\ &\geq C \sum_{n=1}^\infty \alpha_n^{-t} \int_{\alpha_{n+1}}^{\alpha_n} t\alpha^{t-1} d\alpha = C \sum_{n=1}^\infty \alpha_n^{-t}(\alpha_n^t - \alpha_{n+1}^t) \\ &= C \sum_{n=1}^\infty \left(1 - \left(\frac{\alpha_{n+1}}{\alpha_n}\right)^t\right) \geq C \sum_{n=1}^\infty 1 - \frac{1}{2^t} = \infty. \end{aligned}$$

- (c) Similarly to (a) we have

$$\int_{\mathbb{R}^d \setminus \Omega_\alpha} |\psi^0(\omega)| d\omega \leq \int_0^\alpha |\Omega_\beta| d\beta = \int_0^\alpha \frac{1}{t} \beta^{1-t} t \beta^{t-1} |\Omega_\beta| d\beta \leq \frac{1}{t} \alpha^{1-t} \|\psi^0\|_t^t.$$

- (d) Clearly,  $|\Omega_\alpha^\delta| = |\Omega_\alpha^\delta \setminus \Omega_{\alpha/2}| + |\Omega_\alpha^\delta \cap \Omega_{\alpha/2}| \leq |\Omega_\alpha^\delta \setminus \Omega_{\alpha/2}| + |\Omega_{\alpha/2}|$ . On the set  $\Omega_\alpha^\delta \setminus \Omega_{\alpha/2}$  we have  $|\psi^0| \leq \frac{\alpha}{2}$  and  $|\psi^\delta| > \alpha$ . Thus,  $|\psi^0 - \psi^\delta|^2 \geq ||\psi^0| - |\psi^\delta||^2 > \frac{\alpha^2}{4}$  and

$$\frac{\alpha^2}{4} |\Omega_\alpha^\delta \setminus \Omega_{\alpha/2}| \leq \int_{\Omega_\alpha^\delta \setminus \Omega_{\alpha/2}} |\psi^0(\omega) - \psi^\delta(\omega)|^2 d\omega \leq \delta^2.$$

- (e) is trivial.

- (f) is very similar to (d). □

Lemma 2.2 provides the ingredients to show convergence of the regularized functions  $\psi_\alpha^\delta$  to  $\psi^0$ .

**Theorem 2.3.** *Let  $\psi^0 \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $\psi^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  with  $\|\psi - \psi^\delta\|_2 \leq \delta$ . Let  $\alpha = \alpha(\delta)$  be an arbitrary parameter choice with*

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \quad \text{for } \delta \rightarrow 0.$$

*Then, for  $\psi_\alpha^\delta = S_\alpha \psi^\delta$  with  $S_\alpha$  given by Equation (5) the following holds true*

$$\|\psi_\alpha^\delta - \psi^0\|_1 \rightarrow 0.$$

*Moreover, if  $\psi \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  for some  $0 < u < 1$ , a parameter choice*

$$\alpha(\delta) = \delta^{\frac{2}{1+u}} \tag{8}$$

implies

$$\|\psi_\alpha^\delta - \psi^0\|_1 = O\left(\delta^{\frac{2u}{1+u}}\right). \quad (9)$$

for  $\delta \rightarrow 0$ .

*Proof.* Let  $\psi^0 \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$ . By splitting up the domain of integration, we obtain

$$\begin{aligned} \|\psi_\alpha^\delta - \psi^0\|_1 &= \int_{\Omega_\alpha^\delta} |\psi^\delta(\omega) - \psi^0(\omega) - \alpha \operatorname{sign} \psi^\delta(\omega)| \, d\omega + \\ &\quad \int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi_\alpha^\delta(\omega) - \psi^0(\omega)| \, d\omega \end{aligned} \quad (10)$$

Using Lemma 2.2 and the Cauchy-Schwarz inequality, the first term becomes

$$\begin{aligned} &\int_{\Omega_\alpha^\delta} |\psi^\delta(\omega) - \psi^0(\omega) - \alpha \operatorname{sign} \psi^\delta(\omega)| \, d\omega \\ &\leq \int_{\Omega_\alpha^\delta} |\psi^\delta(\omega) - \psi^0(\omega)| \, d\omega + \int_{\Omega_\alpha^\delta} \alpha \, d\omega \\ &\leq \sqrt{|\Omega_\alpha^\delta|} \delta + |\Omega_\alpha^\delta| \alpha \\ &\leq \sqrt{|\Omega_{\alpha/2}| + 4\frac{\delta^2}{\alpha^2}} \delta + \left(|\Omega_{\alpha/2}| + 4\frac{\delta^2}{\alpha^2}\right) \alpha \\ &= o\left(\frac{\delta}{\sqrt{\alpha}}\right) + O\left(\frac{\delta^2}{\alpha}\right) + o(1) \rightarrow 0 \end{aligned} \quad (11)$$

for  $\delta \rightarrow 0$ , where we used that  $\alpha(\delta) \rightarrow 0$  and  $\delta^2/\alpha(\delta) \rightarrow 0$  as well as  $|\Omega_\alpha| = o(\alpha^{-1})$ .

To estimate the second term, note that  $\psi_\alpha^\delta = 0$  on  $\mathbb{R}^d \setminus \Omega_\alpha^\delta$ . Also observe that the function  $\psi^0 \chi_{\mathbb{R}^d \setminus \Omega_{2\alpha}}$ , where  $\chi_{\mathbb{R}^d \setminus \Omega_{2\alpha}}$  is the characteristic function of the set  $\mathbb{R}^d \setminus \Omega_{2\alpha}$ , tends to zero point-wise as  $\alpha \rightarrow 0$ , and it is bounded by  $|\psi^0| \in \mathbb{L}^1(\mathbb{R}^d)$ . Thus, Lebesgue's Dominated Convergence Theorem implies  $\int_{\mathbb{R}^d \setminus \Omega_{2\alpha}} \psi^0(\omega) \, d\omega = o(1)$ . Moreover, by Equation (6) we have  $|S_\alpha \psi^0 - \psi_\alpha^\delta| \leq |\psi^0 - \psi^\delta|$  on  $\mathbb{R}^d$ . Again by Cauchy-Schwarz's inequality and Lemma 2.2, the second term now becomes

$$\begin{aligned} &\int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi_\alpha^\delta(\omega) - \psi^0(\omega)| \, d\omega \\ &\leq \int_{\mathbb{R}^d \setminus \Omega_{2\alpha}} |\psi^0(\omega)| \, d\omega + \int_{\Omega_{2\alpha} \setminus \Omega_\alpha^\delta} |\psi^0(\omega) - \psi_\alpha^\delta(\omega)| \, d\omega \\ &\leq o(1) + \int_{\Omega_{2\alpha} \setminus \Omega_\alpha^\delta} |\psi^0(\omega) - S_\alpha \psi^0(\omega)| \, d\omega + \int_{\Omega_{2\alpha} \setminus \Omega_\alpha^\delta} |S_\alpha \psi^0(\omega) - \psi_\alpha^\delta(\omega)| \, d\omega \\ &\leq o(1) + \alpha |\Omega_{2\alpha} \setminus \Omega_\alpha^\delta| + \int_{\Omega_{2\alpha} \setminus \Omega_\alpha^\delta} |\psi^0(\omega) - \psi^\delta(\omega)| \, d\omega \\ &\leq o(1) + \frac{\delta^2}{\alpha} + \sqrt{|\Omega_{2\alpha} \setminus \Omega_\alpha^\delta|} \delta \\ &\leq o(1) + 2\frac{\delta^2}{\alpha} \rightarrow 0 \end{aligned} \quad (12)$$

for  $\delta \rightarrow 0$ . This, together with Equations (10) and (11), shows the first part of the theorem.

Now assume that  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $\alpha(\delta) = \delta^{\frac{2}{1+u}}$  for some  $0 < u < 1$ . Using Lemma 2.2 with  $t = 1 - u$ , the  $o(1)$ -term in (12) can be replaced by  $O(\alpha^u)$ . Furthermore,  $|\Omega_\alpha| = o(\alpha^{u-1})$ . Collecting Equations (10)-(12), we thus obtain

$$\begin{aligned} \|\psi_\alpha^\delta - \psi^0\|_1 &\leq C \frac{\delta^2}{\alpha} + \sqrt{|\Omega_{\alpha/2}|} \delta + |\Omega_{\alpha/2}| \alpha + O(\alpha^u) \\ &\leq C \left( \frac{\delta^2}{\alpha} + \frac{\delta}{\alpha^{(1-u)/2}} + \alpha^u \right). \end{aligned}$$

The optimal rate is obtained for the given  $\alpha = \delta^{2/(1+u)}$ , where

$$\|\psi_\alpha^\delta - \psi^0\|_1 \leq C \delta^{2u/(1+u)}.$$

□

## 2.2 The discrepancy principle as a-posteriori rule

The a-priori parameter choice rule (8) requires knowledge of some  $0 < u < 1$  for which  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  to achieve the convergence rate (9). However, this information might not be available in practice. In the following, we will show that Morozov's discrepancy principle is an a-posteriori parameter choice rule that guarantees the same convergence rates without requiring knowledge of  $u$ .

For a constant  $\tau > 1$ , the regularization parameter defined via Morozov's discrepancy principle is given by

$$\alpha(\delta, \psi^\delta) := \sup\{\alpha > 0 \mid \|S_\alpha \psi^\delta - \psi^\delta\|_2 \leq \tau \delta\}. \quad (13)$$

We will briefly check under what conditions  $\alpha(\delta, \psi^\delta)$  is finite. The function

$$G(\alpha) := \|S_\alpha \psi^\delta - \psi^\delta\|_2$$

is non-decreasing on  $[0, \infty)$  since  $|S_\alpha \psi^\delta - \psi^\delta|$  is non-decreasing in  $\alpha$  in a point-wise manner. Moreover,  $G$  is continuous on  $[0, \infty)$ . To see this, let  $\alpha \rightarrow \alpha_0 \geq 0$ . Then, by Lebesgue's Dominated Convergence Theorem we have

$$|G(\alpha) - G(\alpha_0)| \leq \|S_\alpha \psi^\delta - S_{\alpha_0} \psi^\delta\|_2 \rightarrow 0$$

since  $S_\alpha \psi^\delta \rightarrow S_{\alpha_0} \psi^\delta$  point-wise and  $|S_\alpha \psi^\delta| \leq |\psi^\delta|$ . Similarly, for  $\alpha \rightarrow \infty$  the function  $S_\alpha \psi^\delta$  tends to zero almost everywhere and thus  $S_\alpha \psi^\delta \rightarrow 0$  in  $\mathbb{L}^2(\mathbb{R}^d)$ . Hence,  $G(\alpha) \rightarrow \|\psi^\delta\|_2$  for  $\alpha \rightarrow \infty$ .

To summarize,  $G$  is continuous and non-decreasing with  $G(0) = 0$  and  $G(\alpha) \rightarrow \|\psi^\delta\|_2$  for  $\alpha \rightarrow \infty$ . Thus, the supremum in (13) is attained as long as  $\|\psi^\delta\|_2 > \tau \delta$  and we have

$$\|S_{\alpha(\delta, \psi^\delta)} \psi^\delta - \psi^\delta\|_2 = \tau \delta.$$

The condition  $\|\psi^\delta\|_2 > \tau\delta$  is known as *signal to noise ratio condition* and states that the energy contained in the “signal”  $\psi^\delta$  is greater than the noise level  $\delta$  resp.  $\tau\delta$  in our case. Note that for sufficiently small  $\delta$  we eventually have  $\|\psi^\delta\|_2 \geq \|\psi^0\|_2 - \|\psi^\delta - \psi^0\|_2 > \tau\delta$  unless  $\psi^0 = 0$ .

The following lemma describes the behavior of  $\alpha$  for  $\delta \rightarrow 0$ .

**Lemma 2.4.** *Let  $\psi^0, \psi^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  with  $\|\psi^0 - \psi^\delta\|_2 \leq \delta$ . Let  $\alpha = \alpha(\delta, \psi^\delta)$  be chosen according to the discrepancy principle (13) with arbitrary  $\tau > 1$ . If  $\psi^0 \neq 0$ , we have*

$$\alpha(\delta, \psi^\delta) \rightarrow 0 \tag{14}$$

for  $\delta \rightarrow 0$ . If  $\psi^0 = 0$ , then

$$\alpha(\delta, \psi^\delta) = \infty \tag{15}$$

holds true for all  $\delta$ .

*Proof.* If  $\psi^0 = 0$ , then  $\alpha(\delta, \psi^\delta) = \infty$  for all  $\delta > 0$  since

$$\|S_\alpha \psi^\delta - \psi^\delta\|_2 \leq \|\psi^\delta\|_2 = \|\psi^\delta - \psi^0\|_2 \leq \delta < \tau\delta$$

holds true for all  $\alpha > 0$ . Now, assume that  $\psi^0 \neq 0$ . As seen before, we then have  $\alpha(\delta, \psi^\delta) < \infty$  for  $\delta$  sufficiently small. If there were a subsequence  $\delta_n \rightarrow 0$  with  $\alpha_n = \alpha(\delta_n, \psi^{\delta_n}) \rightarrow \infty$ , we would obtain

$$\tau\delta_n = \|S_{\alpha_n} \psi^{\delta_n} - \psi^{\delta_n}\|_2 \geq \|S_{\alpha_n} \psi^{\delta_n} - \psi^0\|_2 - \|\psi^0 - \psi^{\delta_n}\|_2 \rightarrow \|\psi^0\|_2.$$

This would contradict  $\psi^0 \neq 0$ , so  $\alpha(\delta, \psi^\delta)$  is bounded for  $\delta$  sufficiently small. Equation (7) now implies

$$\begin{aligned} \|S_\alpha \psi^0 - \psi^0\|_2 &\leq \|S_\alpha \psi^\delta - \psi^\delta\|_2 + \|(\psi^\delta - S_\alpha \psi^\delta) - (\psi^0 - S_\alpha \psi^0)\|_2 \\ &\leq \|S_\alpha \psi^\delta - \psi^\delta\|_2 + \|\psi^\delta - \psi^0\|_2 \\ &\leq (\tau + 1)\delta. \end{aligned} \tag{16}$$

We want to show that  $\alpha \rightarrow 0$  as  $\delta \rightarrow 0$ . Assume the contrary. Then, there exists a sequence  $\delta_n \rightarrow 0$ , but  $\alpha_n > C > 0$ . Since  $(\alpha_n)$  must be bounded, we can w.l.o.g. assume (by extracting a subsequence if necessary) that  $\alpha_n \rightarrow \alpha_0 > 0$ . Therefore, Equation (16) yields

$$\|S_{\alpha_0} \psi^0 - \psi^0\|_2 = \lim_{n \rightarrow \infty} \|S_{\alpha_n} \psi^0 - \psi^0\|_2 \leq \lim_{n \rightarrow \infty} (\tau + 1)\delta_n = 0,$$

which implies  $\psi^0 = 0$  because of  $\alpha_0 > 0$ . This contradicts our assumption on  $\psi^0$ . Hence,

$$\alpha(\delta, \psi^\delta) \rightarrow 0 \text{ for } \delta \rightarrow 0. \tag{17}$$

□

With the preliminary work done so far, we are now able to show that the discrepancy principle yields an a-posteriori parameter choice rule that does *not* saturate.

**Theorem 2.5.** *Let  $\psi^0 \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $\psi^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  with  $\|\psi^0 - \psi^\delta\|_2 \leq \delta$ . Let  $\alpha = \alpha(\delta, \psi^\delta)$  be chosen according to the discrepancy principle (13) with arbitrary  $\tau > 1$ . Then the following holds true*

$$\|\psi_\alpha^\delta - \psi^0\|_1 \rightarrow 0.$$

Moreover, if  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  for some  $0 < u < 1$  we have

$$\|\psi_\alpha^\delta - \psi^0\|_1 = O\left(\delta^{\frac{2u}{1+u}}\right).$$

for  $\delta \rightarrow 0$ .

*Proof.* As seen before,  $\psi^0 = 0$  implies  $S_\alpha \psi^\delta = 0$  for all  $\delta$  and the statements hold trivially. If  $\psi^0 \neq 0$ , we eventually have  $\|S_\alpha \psi^\delta - \psi^\delta\|_2 = \tau\delta$  for  $\delta$  small enough and Lemma 2.4 implies  $\alpha(\delta, \psi^\delta) \rightarrow 0$  for  $\delta \rightarrow 0$ . Using Equation (7) we obtain

$$\begin{aligned} \|S_\alpha \psi^0 - \psi^0\|_2 &\geq \|S_\alpha \psi^\delta - \psi^\delta\|_2 - \|(\psi^\delta - S_\alpha \psi^\delta) - (\psi^0 - S_\alpha \psi^0)\|_2 \\ &\geq \|S_\alpha \psi^\delta - \psi^\delta\|_2 - \|\psi^\delta - \psi^0\|_2 \\ &\geq (\tau - 1)\delta. \end{aligned} \quad (18)$$

Therefore, splitting the domain of integration of the norm into  $\Omega_\alpha$  and  $\mathbb{R}^d \setminus \Omega_\alpha$  yields by using Lemma 2.2 and the fact that  $\psi^0 \in \mathbb{L}^t(\mathbb{R}^d)$  for  $t = 1 - u$

$$\begin{aligned} (\tau - 1)^2 \delta^2 &\leq \|S_\alpha \psi^0 - \psi^0\|_2^2 \\ &= \alpha^2 |\Omega_\alpha| + \int_{\mathbb{R}^d \setminus \Omega_\alpha} |\psi^0(\omega)|^2 d\omega \\ &\leq \alpha^2 |\Omega_\alpha| + \alpha \int_{\mathbb{R}^d \setminus \Omega_\alpha} |\psi^0(\omega)| d\omega \\ &= \alpha^{2-t} \cdot \begin{cases} o(1) & \text{for } t = 1 \\ O(1) & \text{for } t < 1 \end{cases} \end{aligned} \quad (19)$$

as  $\delta \rightarrow 0$ . For  $t = 1$ , Equation (19) shows that  $\delta^2/\alpha \rightarrow 0$  as  $\delta \rightarrow 0$ . For  $t < 1$  we obtain

$$\frac{\delta^2}{\alpha} = \left(\frac{\delta^{4-2t}}{\alpha^{2-t}}\right)^{\frac{1}{2-t}} = \left(\frac{\delta^2}{\alpha^{2-t}}\right)^{\frac{1}{2-t}} \delta^{\frac{2-2t}{2-t}} = O\left(\delta^{\frac{2-2t}{2-t}}\right). \quad (20)$$

Together with Equation (20), the discrepancy principle is a parameter choice fulfilling the assumptions of the first part of Theorem 2.3, and we obtain

$$\|S_\alpha \psi^\delta - \psi^0\|_1 \rightarrow 0.$$

This shows the first part of the theorem. Let now  $\psi^0 \in \mathbb{L}^t(\mathbb{R}^d)$  with  $0 < t < 1$ . In analogy to the derivation of Equations (10) and (11) we get

$$\|S_\alpha \psi^\delta - \psi^0\|_1 \leq \sqrt{|\Omega_\alpha^\delta|} \delta + |\Omega_\alpha^\delta| \alpha + \int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)| d\omega. \quad (21)$$

Due to the discrepancy principle,

$$\tau^2 \delta^2 = \|S_\alpha \psi^\delta - \psi^\delta\|_2^2 \geq \alpha^2 |\Omega_\alpha^\delta|$$

and hence

$$|\Omega_\alpha^\delta| \leq \tau^2 \frac{\delta^2}{\alpha^2}. \quad (22)$$

By setting  $p := 2 - t$  and  $q := \frac{2-t}{1-t}$ , the last term in Equation (21) can be written as

$$\int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)| \, d\omega = \int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)|^{\frac{t}{p}} |\psi^0(\omega)|^{\frac{2}{q}} \, d\omega.$$

Note that  $1/p + 1/q = 1$  and  $1 < p, q < \infty$ . Therefore, Hölder's inequality applies and yields

$$\begin{aligned} \int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)| \, d\omega &\leq \left( \int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)|^t \, d\omega \right)^{\frac{1}{p}} \left( \int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)|^2 \, d\omega \right)^{\frac{1}{q}} \\ &\leq \|\psi^0\|_t^{\frac{t}{2-t}} \|\psi^0 - S_\alpha \psi^\delta\|_2^{\frac{2-2t}{2-t}} \end{aligned} \quad (23)$$

Note that  $S_\alpha \psi^\delta = 0$  on  $\mathbb{R}^d \setminus \Omega_\alpha^\delta$ . Furthermore,

$$\|\psi^0 - S_\alpha \psi^\delta\|_2 \leq \|\psi^0 - \psi^\delta\|_2 + \|\psi^\delta - S_\alpha \psi^\delta\|_2 \leq \delta + \tau \delta = (1 + \tau) \delta.$$

Plugging this into Equation (23) implies

$$\int_{\mathbb{R}^d \setminus \Omega_\alpha^\delta} |\psi^0(\omega)| \, d\omega = O\left(\delta^{\frac{2-2t}{2-t}}\right) \quad (24)$$

Combining Equations (24), (22) and (20) with (21) finally yields

$$\|S_\alpha \psi^\delta - \psi^0\|_1 \leq C \left( \frac{\delta^2}{\alpha} + \delta^{\frac{2-2t}{2-t}} \right) \leq C \delta^{\frac{2-2t}{2-t}} = C \delta^{\frac{2u}{1+u}}$$

with  $u = 1 - t$ . This finishes the proof.  $\square$

**Remark 2.6.** *The convergence results presented in Theorems 2.3 and 2.5 show that continuous soft-shrinkage is a regularization method that does not saturate. This means that the rate, at which the error  $\|S_\alpha \psi^\delta - \psi^0\|_1$  vanishes, can be made arbitrarily close to  $O(\delta)$  by requiring that  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  for  $u$  close to 1. As for linear regularization methods that do not saturate, the discrepancy principle implies the same convergence rates as the a-priori rules. This is not true for regularization methods that saturate, e.g. Tikhonov regularization of linear operators in Hilbert spaces [28]. In the next section, we will show that the convergence rates obtained above are even of optimal order.*

### 3 Order optimality

As we will see in this Section, the convergence rate  $O\left(\delta^{\frac{2u}{1+u}}\right)$  is optimal for the given source representation  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$ . That is, without any further assumption on  $\psi^0$ , one cannot expect a better convergence rate as  $\delta$  vanishes, no matter which regularization method and which parameter choice rule is considered. Many of the basic ideas for the proof, such as usage of the modulus of continuity  $\Omega$  and the worst-case error  $\Delta$ , can be found in more detail in Section 3.2 of [28].

For this section, let us introduce the notation  $X := \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $Y := \mathbb{L}^2(\mathbb{R}^d)$  with  $\|\psi\|_X = \|\psi\|_1 + \|\psi\|_2$  and the standard  $\mathbb{L}^2$ -norm on  $Y$ . Consider the embedding  $T : X \hookrightarrow Y$ . In fact, the soft shrinkage operator considered above is a regularization for  $T$ . For a set  $\mathcal{M} \subset X$  and some  $\delta > 0$ , the modulus of continuity  $\Delta(\delta, \mathcal{M})$  is defined as

$$\Omega(\delta, \mathcal{M}) := \sup\{\|\psi\|_X \mid \psi \in \mathcal{M}, \|T\psi\|_Y \leq \delta\}.$$

Any method for approximating the solution of the ill-posed equation  $T\psi = \psi^\delta$  can be written as an operator  $R : Y \rightarrow X$ . The worst-case error under the information  $\|\psi^0 - \psi^\delta\|_Y \leq \delta$  and  $\psi^0 \in \mathcal{M}$  is given by

$$\Delta(\delta, \mathcal{M}, R) := \sup\{\|R\psi^\delta - \psi\|_X \mid \psi \in \mathcal{M}, \psi^\delta \in Y, \|T\psi - \psi^\delta\|_Y \leq \delta\}.$$

A direct implication is the following

**Lemma 3.1.** *Let  $\mathcal{M} \subset X$ ,  $\delta > 0$ ,  $R : Y \rightarrow X$  be an arbitrary map with  $R(0) = 0$ . Then*

$$\Delta(\delta, \mathcal{M}, R) \geq \Omega(\delta, \mathcal{M}). \quad (25)$$

*Proof.* See Proposition 3.10 in [28].  $\square$

Equation (25) can be used to estimate the worst-case error for *all* regularization methods satisfying  $R(0) = 0$ , which is a very weak restriction. We are particularly interested in lower bounds on  $\Omega(\delta, \mathcal{M})$  for the source sets

$$\mathcal{M} = \mathcal{M}(u, \rho) := \{\psi \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d) \mid \|\psi\|_{1-u} \leq \rho\}. \quad (26)$$

The value of  $\Omega(\delta, \mathcal{M})$  is computed in the following

**Theorem 3.2.** *For  $0 < u < 1$  and  $\rho > 0$ , let the source set  $\mathcal{M}$  be defined as above. Then, the modulus of continuity  $\Omega(\delta, \mathcal{M})$  is given by*

$$\Omega(\delta, \mathcal{M}) = \delta^{\frac{2u}{1+u}} \rho^{\frac{1-u}{1+u}} + \delta. \quad (27)$$

*Proof.* Let  $\psi \in \mathcal{M}$  with  $\|T\psi\|_Y = \|\psi\|_2 \leq \delta$ . By Hölder's inequality, we obtain similarly to the proof of Theorem 2.5 that

$$\begin{aligned} \|\psi\|_1 + \|\psi\|_2 &= \int_{\mathbb{R}^d} |\psi(\omega)|^{\frac{2u}{1+u}} |\psi(\omega)|^{\frac{1-u}{1+u}} d\omega + \|\psi\|_2 \\ &\leq \|\psi\|_2^{\frac{2u}{1+u}} \|\psi\|_{1-u}^{\frac{1-u}{1+u}} + \delta \leq \delta^{\frac{2u}{1+u}} \rho^{\frac{1-u}{1+u}} + \delta. \end{aligned}$$

By taking the supremum over all possible  $\psi$ , we get

$$\Omega(\delta, \mathcal{M}) \leq \delta^{\frac{2u}{1+u}} \rho^{\frac{1-u}{1+u}} + \delta.$$

To see that this estimate is sharp, we construct a  $\psi$  for which the above inequality chain is sharp. Consider the function

$$\bar{\psi} := \delta^{\frac{2}{1+u}} \rho^{-\frac{1-u}{1+u}} \chi_V,$$

where  $\chi_V$  is the characteristic function on an arbitrary measurable set  $V \subset \mathbb{R}^d$  with measure

$$|V| = \delta^{-2\frac{1-u}{1+u}} \rho^{2\frac{1-u}{1+u}}.$$

Since for all  $0 < p < \infty$  we have  $\|\bar{\psi}\|_p = \delta^{\frac{2}{1+u}} \rho^{-\frac{1-u}{1+u}} |V|^{\frac{1}{p}}$ , one easily sees that

$$\|\bar{\psi}\|_{1-u} = \rho, \quad \|\bar{\psi}\|_1 = \delta^{\frac{2u}{1+u}} \rho^{\frac{1-u}{1+u}} \quad \text{and} \quad \|\bar{\psi}\|_2 = \delta.$$

Thus,  $\bar{\psi} \in \mathcal{M}$  with  $\|\bar{\psi}\|_2 = \delta$  and

$$\Omega(\delta, \mathcal{M}) \geq \|\bar{\psi}\|_1 + \|\bar{\psi}\|_2 = \delta^{\frac{2u}{1+u}} \rho^{\frac{1-u}{1+u}} + \delta.$$

□

Equation (27) combined with Equation (25) now yield the worst-case estimate

$$\Delta(\delta, \mathcal{M}, R) \geq \delta^{\frac{2u}{1+u}} \rho^{\frac{1-u}{1+u}} + \delta.$$

for an arbitrary regularization method  $R$ . In particular, the convergence rate for  $\delta \rightarrow 0$  cannot be better than  $O(\delta^{\frac{2u}{1+u}})$ , so that the soft shrinkage operator  $S_\alpha$  discussed in the previous section is an order optimal regularization method.

## 4 Shrinkage + Isometries

The convergence and convergence rate results obtained for the considered parameter choice rules for the functional  $F_\alpha$  can be easily generalized to functionals that emerge from suitable transformations  $A$ . In this section, we make the following assumptions:

$$A : X \rightarrow \mathbb{L}^2(\mathbb{R}^d) \quad \text{is linear} \quad (28)$$

$$\|\varphi\|_X = \|A(\varphi)\|_{L^2} \quad (29)$$

$$S_\alpha(A(\varphi)) \in R(A) \text{ for all } \varphi \in X. \quad (30)$$

Here,  $X$  denotes a Banach space and  $S_\alpha$  is the shrinkage operator defined in (5). Condition (29) implies the injectivity of  $A$ , so that by Condition (30) the shrunk function  $S_\alpha(A(\varphi))$  can be mapped back one-to-one to  $X$ . The importance of this range condition has also been pointed out in [24]. We define the spaces  $\mathcal{Y}^t$  by

$$\varphi \in \mathcal{Y}^t \Leftrightarrow \|\varphi\|_{\mathcal{Y}^t} < \infty,$$

where  $\|\cdot\|_{\mathcal{Y}^t}$  is defined by

$$\|\varphi\|_{\mathcal{Y}^t} := \|A(\varphi)\|_{\mathbb{L}^t(\mathbb{R}^d)}. \quad (31)$$

In particular, we have  $\mathcal{Y}^2 = X$ . Please note that  $\|\cdot\|_{\mathcal{Y}^t}$  is for  $t < 1$  not a norm anymore. The functional under consideration is now defined by

$$J_\alpha(\varphi) = \|\varphi^\delta - \varphi\|_X^2 + \|\varphi\|_{\mathcal{Y}^1}. \quad (32)$$

The main goal of this section is to show that the results of Section 2 carry over to the above setting. Some applications will be presented in the following section. For given data  $\varphi^\delta, \varphi^0$ , we define the associated functions

$$\psi^\delta = A(\varphi^\delta), \quad \psi^0 = A(\varphi^0) \quad (33)$$

and observe

$$\|\psi^\delta - \psi^0\|_{\mathbb{L}^t(\mathbb{R}^d)} = \|\varphi^\delta - \varphi^0\|_{\mathcal{Y}^t}. \quad (34)$$

Let us start with an observation on the minimizers of the functional (32):

**Proposition 4.1.** *The minimizer  $\varphi_\alpha^\delta$  of (32) is computed by*

$$\varphi_\alpha^\delta = A^{-1}S_\alpha(A(\varphi^\delta)) := R_\alpha(\varphi^\delta). \quad (35)$$

*Proof.* Because of (29), (30),  $\varphi_\alpha^\delta$  is well defined. The (unique) minimizer of  $F_\alpha$  with data  $\psi^\delta$  is computed by  $S_\alpha(\psi^\delta) = S_\alpha(A(\varphi^\delta))$ , see (5). Now let us assume that  $\varphi_\alpha^\delta = A^{-1}S_\alpha(A(\varphi^\delta))$  is not a minimizer of  $J_\alpha$ . Then there exists  $\tilde{\varphi}$  with  $J_\alpha(\tilde{\varphi}) < J_\alpha(\varphi_\alpha^\delta)$ . It follows

$$\begin{aligned} F_\alpha(A(\tilde{\varphi})) &= \|A(\tilde{\varphi}) - A(\varphi^\delta)\|_2^2 + \alpha\|A(\tilde{\varphi})\|_1 \\ &\stackrel{(34)}{=} \|\tilde{\varphi} - \varphi^\delta\|_X^2 + \alpha\|\tilde{\varphi}\|_{\mathcal{Y}^1} = J_\alpha(\tilde{\varphi}) \\ &< \|\varphi^\delta - \varphi_\alpha^\delta\|_X^2 + \alpha\|\varphi_\alpha^\delta\|_{\mathcal{Y}^1} \\ &\stackrel{(34)}{=} \|A(\varphi^\delta) - A(\varphi_\alpha^\delta)\|_2^2 + \alpha\|A(\varphi_\alpha^\delta)\|_1 \\ &= \|\psi^\delta - S_\alpha(\psi^\delta)\|_2^2 + \alpha\|S_\alpha(\psi^\delta)\|_1 = F_\alpha(S_\alpha(\psi^\delta)), \end{aligned}$$

which is in contrast to the fact that  $S_\alpha(\psi^\delta)$  is the minimizer of  $F_\alpha$ . Now assume there exists another different minimizer  $\tilde{\varphi}$  of  $J_\alpha$ , then, due to the injectivity of  $A$ ,  $A(\tilde{\varphi})$  would be a minimizer of  $F_\alpha$  different from  $S_\alpha(\psi^\delta)$ , which is in contrast to the fact that  $F_\alpha$  has a unique minimizer.  $\square$

From the isometries of the norm follows in particular for the data error

$$\|\varphi^\delta - \varphi^0\| \leq \delta \implies \|\psi^\delta - \psi^0\| \leq \delta.$$

We will also use Morozov's discrepancy principle for the computation of an optimal approximation to  $\varphi^\delta$ , i.e. we choose

$$\begin{aligned} \alpha_\delta(\delta, \varphi^\delta) &= \sup\{\alpha > 0 \mid \|R_\alpha(\varphi^\delta) - \varphi^\delta\|_X \leq \tau\delta\} \\ &= \sup\{\alpha > 0 \mid \|S_\alpha(\psi^\delta) - \psi^\delta\|_2 \leq \tau\delta\}. \end{aligned} \quad (36)$$

Therefore, Morozov's discrepancy principle applied to  $F_\alpha$  and  $J_\alpha$  yields the same regularization parameter. In particular, Lemma 2.4 holds also for Morozov's discrepancy principle for  $J_\alpha$ . Before investigating the discrepancy principle in more detail, we will first give a result for an a priori parameter choice rule:

**Theorem 4.2.** Let  $\varphi^0 \in \mathcal{Y}^1 \cap X$  and  $\varphi^\delta \in X$  with  $\|\varphi^0 - \varphi^\delta\|_X \leq \delta$ . Let  $\alpha = \alpha(\delta)$  be an arbitrary parameter choice with

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \quad \text{for } \delta \rightarrow 0.$$

Then, for  $\varphi_\alpha^\delta = R_\alpha \varphi^\delta$  with  $R_\alpha$  given by Equation (35) the following holds true

$$\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} \rightarrow 0.$$

Moreover, if  $\varphi^0 \in \mathcal{Y}^{1-u}$  for some  $0 < u < 1$ , a parameter choice

$$\alpha(\delta) = \delta^{\frac{2}{1+u}} \tag{37}$$

implies

$$\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} = O\left(\delta^{\frac{2u}{1+u}}\right). \tag{38}$$

for  $\delta \rightarrow 0$ .

*Proof.* The proof is straight forward and based on Theorem 2.3. First, we conclude also  $\|\psi^0 - \psi^\delta\|_2 \leq \delta$ . For the minimizers  $\varphi_\alpha^\delta$  and  $\psi_\alpha^\delta$  of the functionals  $J_\alpha$  and  $F_\alpha$ , resp., follows

$$\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} = \|\psi_\alpha^\delta - \psi^0\|_1 \xrightarrow{(9)} 0. \tag{39}$$

If we have  $\psi^0 \in \mathcal{Y}^t$ , so follows in particular also  $\|\psi^0\|_t = \|\varphi^0\|_{\mathcal{Y}^t} < \infty$ , i.e.  $\psi^0 \in \mathbb{L}^t(\mathbb{R}^d)$ , and by the norm isometry and (9) follows

$$\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} = \|\psi_\alpha^\delta - \psi^0\|_1 = O\left(\delta^{\frac{2u}{1+u}}\right). \tag{40}$$

for the parameter choice rule (37).  $\square$

Let us now come to the convergence result for Morozov's discrepancy principle:

**Theorem 4.3.** Let  $\varphi^0 \in \mathcal{Y}^1 \cap X$  and  $\varphi^\delta \in X$  with  $\|\varphi^0 - \varphi^\delta\|_X \leq \delta$ . Let  $\alpha = \alpha(\delta, \psi^\delta)$  be chosen according to the discrepancy principle (36) with arbitrary  $\tau > 1$ . Then the following holds true

$$\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} \rightarrow 0.$$

Moreover, if  $\varphi^0 \in \mathcal{Y}^{1-u}$  for some  $0 < u < 1$  we have

$$\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} = O\left(\delta^{\frac{2u}{1+u}}\right).$$

for  $\delta \rightarrow 0$ .

*Proof.* According to (36), Morozov's discrepancy principle for  $F_\alpha$  and  $J_\alpha$  yields the same regularization parameter for each data error level  $\delta$ , and thus the proof follows by the norm isometry  $\|\varphi_\alpha^\delta - \varphi^0\|_{\mathcal{Y}^1} = \|\psi_\alpha^\delta - \psi^0\|_{\mathbb{L}^1(\mathbb{R}^d)}$  from Theorem 2.5.  $\square$

## 5 Continuous Fourier soft-shrinkage

As an example of an isometry discussed in the last section consider the Fourier transform  $A = \mathcal{F}$ . Clearly,  $\mathcal{F}$  is an isometry on  $\mathbb{L}^2(\mathbb{R}^d)$ . The scale of spaces  $\mathcal{Y}^t$  is given by

$$\mathcal{Y}^t := \{g : \mathbb{R}^d \rightarrow \mathbb{R} \mid \hat{g} \in \mathbb{L}^t(\mathbb{R}^d)\}$$

for  $0 < t \leq 1$ . Since  $\mathcal{F}(X) = X = \mathbb{L}^2(\mathbb{R}^d)$ , the relation  $S_\alpha(\mathcal{F}(X)) \subset \mathcal{F}(X)$  trivially holds. The operator  $R_\alpha$ , which we call the *Fourier soft-shrinkage operator* here, is given by

$$R_\alpha = \mathcal{F}^{-1} \circ S_\alpha \circ \mathcal{F}.$$

We have

**Theorem 5.1.** *Let  $g^0 \in \mathcal{Y}^1 \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $g^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  with  $\|g - g^\delta\|_2 \leq \delta$ . Furthermore, let  $\alpha = \alpha(\delta)$  be an arbitrary parameter choice rule with*

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \quad \text{for } \delta \rightarrow 0.$$

Then we have for  $\delta \rightarrow 0$

$$\|R_\alpha g^\delta - g^0\|_{\mathcal{Y}^1} \rightarrow 0. \quad (41)$$

Moreover, if  $g^0 \in \mathcal{Y}^{1-u}$  for some  $0 < u < 1$ , a parameter choice

$$\alpha(\delta) = \delta^{\frac{2}{1+u}}$$

yields

$$\|R_\alpha g^\delta - g^0\|_{\mathcal{Y}^1} = O\left(\delta^{\frac{2u}{u+1}}\right). \quad (42)$$

for  $\delta \rightarrow 0$ . The same results hold true if the parameter  $\alpha$  is chosen according to the discrepancy principle

$$\alpha(\delta, g^\delta) := \sup\{\alpha > 0 \mid \|R_\alpha g^\delta - g^\delta\|_2 \leq \tau\delta\} \quad (43)$$

with arbitrary  $\tau > 1$ .

### 5.1 Discussion of $\mathcal{Y}^t$ -spaces

The ‘‘smoothness’’ condition  $g^0 \in \mathcal{Y}^t$  resp.  $\hat{g}^0 \in \mathbb{L}^t(\mathbb{R}^d)$  with  $t < 1$  is somewhat non-standard. Note that  $\mathbb{L}^t(\mathbb{R}^d)$  is only a quasi-normed space since the map  $g \mapsto \|\hat{g}\|_t$  does not fulfill the triangle inequality. We want to give an interpretation of these spaces by showing that the Sobolev spaces  $H^s(\mathbb{R}^d)$  are contained in  $\mathcal{Y}^t$  for properly chosen numbers  $s$  and  $t$ .

**Theorem 5.2** ( $H^s(\mathbb{R}^d) \hookrightarrow \mathcal{Y}^t$  Embedding Theorem).  
For  $s > 0$  and  $t_0 := \frac{2d}{d+2s} < t \leq 2$  the inclusion  $H^s(\mathbb{R}^d) \subset \mathcal{Y}^t$  holds.

*Proof.* Let  $g \in H^s(\mathbb{R}^d)$ . For  $\alpha \geq 0$ , let  $\Omega_\alpha$  be the set

$$\Omega_\alpha = \{\omega \in \mathbb{R}^d \mid |\hat{g}(\omega)| > \alpha\}.$$

First, we obtain by Fubini's Theorem

$$\begin{aligned}\|g\|_{H^s}^2 &= \int_{\mathbb{R}^d} \Delta(\omega)^s |\hat{g}(\omega)|^2 d\omega = \int_{\mathbb{R}^d} \Delta(\omega)^s \int_0^{|\hat{g}(\omega)|} 2\alpha d\alpha d\omega \\ &= \int_0^\infty 2\alpha \int_{\Omega_\alpha} \Delta(\omega)^s d\omega d\alpha.\end{aligned}\quad (44)$$

Note that the integrant  $\Delta(\omega) = (1 + |\omega|^2)^s$  is monotonically increasing in  $|\omega|$ . Thus, amongst all possible measurable sets  $M$  with  $|M| = |\Omega_\alpha|$ , the inner integral achieves its minimal value if  $M$  is a ball centered around  $\omega = 0$ . Since the volume of a  $d$ -dimensional ball with radius  $R$  is given by  $\frac{\omega_d}{d} R^d$ , where  $\omega_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$  denotes the area of the  $d$ -dimensional unit sphere,  $R$  can be computed as

$$R = \left( \frac{d|\Omega_\alpha|}{\omega_d} \right)^{1/d}.$$

By using polar coordinates, the inner integral can thus be estimated as

$$\begin{aligned}\int_{\Omega_\alpha} \Delta(\omega)^s d\omega &\geq \int_M \Delta(\omega)^s d\omega = \int_0^R (1 + r^2)^s \omega_d r^{d-1} dr \\ &\geq \int_0^R r^{2s+d-1} \omega_d dr = \frac{\omega_d}{2s+d} R^{2s+d} \\ &= C |\Omega_\alpha|^{\frac{2s+d}{d}},\end{aligned}$$

where  $C$  only depends on  $s$  and the dimension  $d$ . If we plug this into Equation (44), we obtain

$$\|g\|_{H^s}^2 \geq C \int_0^\infty 2\alpha |\Omega_\alpha|^{\frac{2s+d}{d}} d\alpha \quad (45)$$

Let  $t_0 < t \leq 2$ . Recall that  $\|\hat{g}\|_2^2 = \int_0^\infty 2\alpha |\Omega_\alpha| d\alpha$ . Using Lemma 2.2(a) and Hölder's inequality for  $p = \frac{2s+d}{2s}$  and  $q = \frac{2s+d}{d}$ , we therefore obtain

$$\begin{aligned}\frac{1}{t} \|\hat{g}\|_t^t &= \int_0^\infty \alpha^{t-1} |\Omega_\alpha| d\alpha \\ &\leq \int_0^1 \alpha^{t-1} |\Omega_\alpha| d\alpha + \frac{1}{2} \int_1^\infty 2\alpha |\Omega_\alpha| d\alpha \\ &\leq \left( \int_0^1 \alpha^{p(t-1-1/q)} d\alpha \right)^{1/p} \left( \int_0^1 \alpha^{q/q} |\Omega_\alpha|^q d\alpha \right)^{1/q} + \frac{1}{2} \|\hat{g}\|_2^2.\end{aligned}$$

With  $t_0 = \frac{2d}{d+2s} = 2/q$ , the first integral is finite since

$$p \left( t - 1 - \frac{1}{q} \right) > p \left( \frac{1}{q} - 1 \right) = -1.$$

The second integral can be estimated by Equation (45) and we obtain

$$\|\hat{g}\|_t^t \leq C \left( \|g\|_{H^s}^{2/q} + \|\hat{g}\|_2^2 \right) < \infty.$$

with a constant  $C < \infty$  depending on  $s, t$  and  $d$ . Therefore,  $\hat{g} \in \mathbb{L}^t(\mathbb{R}^d)$ .  $\square$

We note that  $t > 2d/(d+2s) = t_0$  in Theorem 5.2 is sharp in the sense that it cannot be extended to  $t \geq t_0$ . We give a one-dimensional counterexample, i.e.  $d = 1$  and  $t_0 = 2/(1+2s)$ . The function  $g$  with

$$\hat{g}(\omega) = \begin{cases} (\omega \ln(\omega))^{-1/t_0} & \text{for } \omega > 2 \\ 0 & \text{else} \end{cases}$$

belongs to  $H^s(\mathbb{R})$ , but not to  $\mathbb{L}^{t_0}(\mathbb{R})$ : With the substitution  $\omega = \exp(y)$  we obtain

$$\|g\|_{H^s(\mathbb{R})}^2 = \int_2^\infty \frac{(1+\omega^2)^s}{(\omega \ln(\omega))^{2/t_0}} d\omega = \int_{\ln(2)}^\infty \frac{(1+\exp(2y))^s}{\exp(y(2s+1))y^{2/t_0}} \exp(y) dy < \infty$$

since  $2/t_0 = 1+2s > 1$ , but

$$\|\hat{g}\|_{t_0}^{t_0} = \int_2^\infty \omega \ln(\omega) d\omega = \int_{\ln(2)}^\infty \exp(-y) \frac{1}{y} \exp(y) dy = \infty.$$

On the other hand, for any  $0 < t < 1$ ,  $\mathcal{Y}^t$  contains functions which need not be differentiable at all. For example, define  $g$  by

$$\hat{g}(\omega) = \begin{cases} n^{-2/t} & \text{if } 2^n \leq \omega < 2^{n+1} \\ 0 & \text{else.} \end{cases} \quad (46)$$

Then, clearly,  $g \in \mathcal{Y}^t$ , but

$$\|g\|_{H^s(\mathbb{R})}^2 = \sum_{n=1}^\infty n^{-4/t} \int_{2^n}^{2^{n+1}} (1+\omega^2)^s d\omega \geq \sum_{n=1}^\infty n^{-4/t} 2^{2ns} = \infty$$

for *any*  $s > 0$ . We have constructed a uniformly continuous function  $g$  which is not contained in any  $H^s(\mathbb{R})$  for  $s > 0$ . See Figure 2 for an illustration of  $g$  defined by Equation (46) with the choice  $t = 0.5$  (left),  $t = 0.75$  (middle) and  $t = 1.0$ .

## 6 Wavelet shrinkage

In image processing, one often considers images as intensity functions  $g$  given on the unit square  $\Omega = [0, 1]^2$ . Denoising such images by Fourier methods leads

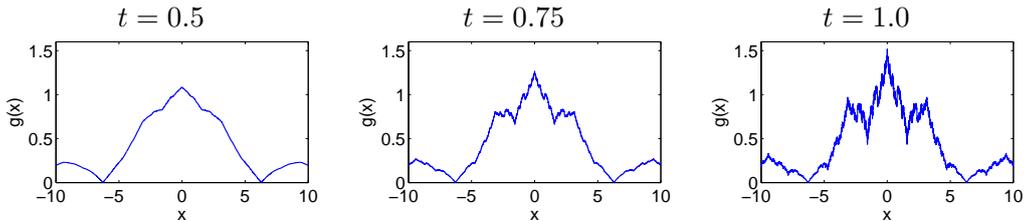


Figure 2: Functions in  $\mathcal{Y}^t$  which are not in  $H^s(\mathbb{R})$  for any  $s > 0$ .

to artifacts visible in the whole domain since the analyzing wave functions have a global support. In wavelet analysis, one uses small-support functions that do not produce global artifacts.

A wavelet basis  $\{\psi_{j,k}^{(d)}\}$  is an orthonormal basis of  $\mathbb{L}^2(\Omega)$  consisting of properly scaled and shifted versions of usually three mother wavelets  $\psi_{0,0}^{(1)}$ ,  $\psi_{0,0}^{(2)}$  and  $\psi_{0,0}^{(3)}$  encoding horizontal, vertical and diagonal information.  $j$  denotes the shift parameter and  $k$  denotes the scaling parameter. For a detailed introduction see for instance [29, 30, 31].

A function  $g \in \mathbb{L}^2(\Omega)$  can be expanded into the wavelet basis by

$$g = \sum_{j,k,d} c_{j,k,d} \psi_{j,k}^{(d)} \quad \text{with} \quad c_{j,k,d} = \langle g, \psi_{j,k}^{(d)} \rangle,$$

so that  $\|g\|_{\mathbb{L}^2(\Omega)} = \|c\|_2$ . The  $\mathbb{L}^2$ -norm can be expressed in terms of the  $\ell^2$ -norm of the wavelet coefficients. For the following, we will denote by

$$\mathcal{W} : \mathbb{L}^2(\Omega) \rightarrow \ell^2 \tag{47}$$

$$g \mapsto c \tag{48}$$

the Wavelet transform, the above norm equality shows that it is an isometry. Similarly, the  $\ell^p$ -norm<sup>1</sup> of the wavelet coefficients can be shown to be equivalent to the norm in certain Besov spaces  $B_{pp}^s(\Omega)$ . We do not want to introduce Besov spaces here, since we will only use the equivalent sequence norm. Assume that  $s - \frac{2}{p} > -2$  and that the wavelets are sufficiently smooth and possess enough vanishing moments. Then the norm equivalence

$$\|g\|_{B_{pp}^s(\Omega)} \asymp \left( \sum_{j,k,\psi} 2^{skp+k(p-2)} |c_{jk\psi}|^p \right)^{1/p} \tag{49}$$

holds [16]. We are especially interested to reconstruct images in  $B_{11}^1(\Omega)$ , since it is close to the space of bounded variation [18, 19, 20]. From Equation (49) the norm in  $B_{11}^1(\Omega)$  turns out to be equivalent to the  $\ell^1$ -norm of the coefficients:

$$\|g\|_{B_{11}^1(\Omega)} \asymp \sum_{j,k,\psi} |c_{jk\psi}| = \|c\|_1.$$

In order to apply our theory, we define the operators

$$I : \ell_2 \rightarrow \mathbb{L}^2(\Omega) \tag{50}$$

$$Ic = \sum_{n=1}^{\infty} c_n \chi_{[n-1,n)} \tag{51}$$

$$A = I \circ \mathcal{W}. \tag{52}$$

The operator  $I$  and thus also  $A$  are invertible on their range, and the operator

$$R_\alpha = A^{-1} S_\alpha A$$

---

<sup>1</sup>To be precise,  $\|\cdot\|_p$  is only a quasi-norm for  $p < 1$ .

describes the well known Wavelet Soft Shrinkage operation. The belonging  $\mathcal{Y}^t$ -spaces are formed by all functions  $g$  with

$$\|g\|_{\mathcal{Y}^t} = \|Ag\|_{\mathbb{L}^t} = \|c\|_{\ell^t} < \infty ,$$

where  $c$  denotes the sequence of Wavelet coefficients of  $g$ . With respect to the Besov norm equivalence (49), these are the Besov spaces  $B_{pp}^s$  where  $skp + k(p-2) = 0$ , i.e.

$$\|g\|_{B_{pp}^s(\Omega)} \asymp \left( \sum_{j,k,\psi} |c_{j k \psi}|^p \right)^{1/p} = \|c\|_p \quad \text{for } 0 < p < 1 \text{ and } s = \frac{2-p}{p}.$$

In particular,  $\mathcal{Y}^1 = B_{11}^1$ , and the following Theorem shows that shrinking the wavelet coefficients leads to a regularization method that converges in  $B_{11}^1(\Omega)$ .

**Theorem 6.1.** *Let  $g^0 \in B_{11}^1(\Omega)$  and  $g^\delta \in \mathbb{L}^2(\Omega)$  with  $\|g - g^\delta\|_{\mathbb{L}^2(\Omega)} \leq \delta$ . Furthermore, let  $\alpha = \alpha(\delta)$  be an arbitrary parameter choice rule with*

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \quad \text{for } \delta \rightarrow 0.$$

Then we have for  $\delta \rightarrow 0$

$$\|R_\alpha g^\delta - g^0\|_{B_{11}^1(\Omega)} \rightarrow 0. \quad (53)$$

Moreover, if  $g^0 \in B_{pp}^s(\Omega)$  for some  $0 < p < 1$  and  $s = \frac{2-p}{p}$ , a parameter choice

$$\alpha(\delta) = \delta^{\frac{1+s}{s}}$$

yields

$$\|R_\alpha g^\delta - g^0\|_{B_{11}^1(\Omega)} = O\left(\delta^{\frac{s-1}{s}}\right). \quad (54)$$

for  $\delta \rightarrow 0$ . The same results hold true if the parameter  $\alpha$  is chosen according to the discrepancy principle

$$\alpha(\delta, g^\delta) := \sup\{\alpha > 0 \mid \|R_\alpha g^\delta - g^\delta\|_{\mathbb{L}^2(\Omega)} \leq \tau \delta\} \quad (55)$$

with arbitrary  $\tau > 1$ .

*Proof.* We check whether the conditions of Theorem 4.2, especially Equations (28)-(30) are fulfilled. Clearly,  $A : X \rightarrow \mathbb{L}^2(\mathbb{R})$  is linear and the canonical embedding  $I$  is an isometry from  $\ell^p$  to  $\mathbb{L}^p(\mathbb{R})$  for every  $0 < p \leq \infty$ , so that

$$\|Ag\|_{\mathbb{L}^2(\mathbb{R})} = \|\mathcal{W}g\|_2 = \|g\|_X$$

for all  $g \in X$ . Moreover,  $Ag$  is a piecewise constant function, so  $S_\alpha(A\varphi)$  is piecewise constant as well and lies in the range of the operator  $I$ . Since the sequence  $I^{-1}S_\alpha(Ag)$  is a shrunked version of the wavelet coefficients of  $g$ , it is an  $\ell^2$ -sequence again and thus belongs to the range of  $\mathcal{W}$ . Therefore,  $S_\alpha(Ag) \in R(A)$ . As shown above, the spaces  $\mathcal{Y}^p$  are given by

$$g \in \mathcal{Y}^p \iff \|Ag\|_{\mathbb{L}^p(\mathbb{R})} = \|\mathcal{W}g\|_p < \infty \iff g \in B_{pp}^s(\Omega)$$

for  $0 < p < 1$  and  $s = \frac{2-p}{p}$ . An application of Theorem 4.2 now shows the result.  $\square$

We wish to conclude this Section with an observation on the reconstruction of functions with *sparse* Wavelet expansions, that is if only a finite number of Wavelet coefficients are nonzero. These functions can be reconstructed with an accuracy of almost  $O(\delta)$ : Indeed, as these functions belong to any Besov space  $B_{pp}^s$ , the source condition in Theorem 6.1 is fulfilled for arbitrary large  $s$ , and we get

**Theorem 6.2.** *Assume that the function  $g^0$  has a sparse representation with respect to a sufficiently smooth Wavelet basis. With the parameter choice rules in Theorem 6.1 the Wavelet Soft Shrinkage  $R_\alpha$  yields a convergence rate of  $O(\delta^{(s-1)/s})$  for all  $s > 1$ .*

If, in a sloppy way, we consider a problem as well posed if a reconstruction with rate  $O(\delta)$  is possible, then a source condition that requires a sparse expansion is so strong that it makes the problem almost well posed.

## 7 Application to blind deconvolution

Blind deconvolution is an image processing problem where both an image  $f$  and a kernel function  $k$  have to be reconstructed from a blurred image  $g^0$  satisfying the convolution equation

$$\mathcal{K}(f, k) = f * k = g^0.$$

Often, even only a noisy version  $g^\delta$  of  $g^0$  is available. We will consider the convolution defined as

$$\mathcal{K} : \mathbb{L}^2(\mathbb{R}^d) \times \mathbb{L}^2(\mathbb{R}^d) \rightarrow \mathbb{L}^\infty(\mathbb{R}^d),$$

where typically  $d = 2$  is taken in image processing problems. Then, the Fourier convolution theorem

$$f * k = g^0 \quad \iff \quad \hat{f}\hat{k} = \hat{g}^0$$

implies that the equation has infinitely many solutions. Thus, it is common to consider the  $(f^\dagger, k^\dagger)$ -minimum-norm-solution (MNS)  $(f^\dagger, k^\dagger)$ , which satisfies

$$\begin{aligned} f^\dagger * k^\dagger &= g^0 \quad \text{and} \\ \|f^\dagger - \bar{f}\|_2^2 + \|k^\dagger - \bar{k}\|_2^2 &= \min_{f * k = g^0} \|f - \bar{f}\|_2^2 + \|k - \bar{k}\|_2^2 \end{aligned}$$

for some a-priori chosen estimator pair  $(f^\dagger, k^\dagger) \in \mathbb{L}^2(\mathbb{R}^d)^2$ . Note that the  $(0, 0)$ -MNS is still not unique [32]. In [33], we choose a normalized, symmetric and non-negative kernel estimator  $\bar{k} \in \mathbb{L}^2(\mathbb{R}^d)$ , i.e.

$$\int_{\mathbb{R}^d} k(x) dx = 1, \quad \bar{k}(-x) = \bar{k}(x) \quad \text{and} \quad k(x) \geq 0,$$

and “link” the image estimator  $\bar{f}$  to  $g^0$  by setting

$$\hat{f}(\omega) := \begin{cases} \hat{g}^0(\omega) & \text{if } \hat{k}(\omega) \geq 0 \\ -\hat{g}^0(\omega) & \text{if } \hat{k}(\omega) < 0 \end{cases}.$$

Then, for each  $g^0 \in \mathcal{Y} \cap \mathbb{L}^2(\mathbb{R}^d)$ , where

$$\mathcal{Y} := \{g : \mathbb{R}^d \rightarrow \mathbb{R} \mid \hat{g} \in \mathbb{L}^1(\mathbb{R}^d)\},$$

there exists a unique  $(\bar{f}, \bar{k})$ -MNS  $(f^\dagger, k^\dagger)$  that can be computed in a fast and non-iterative fashion. The operator

$$L : \mathcal{Y} \cap \mathbb{L}^2(\mathbb{R}^d) \rightarrow \mathbb{L}^2(\mathbb{R}^d) \times \mathbb{L}^2(\mathbb{R}^d), \quad g \mapsto (f^\dagger, k^\dagger)$$

is continuous between those spaces and

$$\|L(g_1) - L(g_2)\|_2^2 \leq C (\|\hat{g}_1 - \hat{g}_2\|_1 + \|\hat{g}_1 - \hat{g}_2\|_2^2). \quad (56)$$

holds for  $g_1, g_2 \in \mathcal{Y} \cap \mathbb{L}^2(\mathbb{R}^d)$  with a constant  $C$  independent of  $g_1$  and  $g_2$ . The inversion operator  $L$  can be thought to be the non-linear equivalent to a linear pseudo-inverse. For proofs of the above statements and algorithms to compute the MNS, we refer to [33].

Noisy data  $g^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  satisfying  $\|g^0 - g^\delta\|_2 \leq \delta$  does not belong to  $\mathcal{Y}$  in general. Thus, the inversion operator  $L$  cannot be applied. However, by means of Fourier soft-shrinkage,  $g^0$  can be stably approximated by  $g^\delta$  in  $\mathcal{Y}$ . With the appropriate “smoothness spaces”

$$\mathcal{Y}^t := \{g : \mathbb{R}^d \rightarrow \mathbb{R} \mid \hat{g} \in \mathbb{L}^t(\mathbb{R}^d)\},$$

we readily have

**Theorem 7.1.** *Let  $R_\alpha$  be the Fourier soft-shrinkage operator defined by  $\widehat{R_\alpha g} = S_\alpha \hat{g}$ . Let  $g^0 \in \mathcal{Y} \cap \mathbb{L}^2(\mathbb{R}^d)$  and  $g^\delta \in \mathbb{L}^2(\mathbb{R}^d)$  with  $\|g^0 - g^\delta\|_2 \leq \delta$ . Furthermore, let  $\alpha = \alpha(\delta)$  be an arbitrary parameter choice rule with*

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \quad \text{for } \delta \rightarrow 0.$$

*Then we have for  $\delta \rightarrow 0$*

$$\|L(R_\alpha g^\delta) - L(g^0)\|_2 \rightarrow 0. \quad (57)$$

*Moreover, if  $g^0 \in \mathcal{Y}^{1-u}$  for some  $0 < u < 1$ , a parameter choice*

$$\alpha(\delta) = \delta^{\frac{2}{1+u}}$$

*yields*

$$\|L(R_\alpha g^\delta) - L(g^0)\|_2 = O\left(\delta^{\frac{u}{u+1}}\right). \quad (58)$$

*for  $\delta \rightarrow 0$ . The same results hold true if the parameter  $\alpha$  is chosen a-posteriori according to the discrepancy principle*

$$\alpha(\delta, g^\delta) := \sup\{\alpha > 0 \mid \|R_\alpha g^\delta - g^\delta\|_2 \leq \tau \delta\} \quad (59)$$

*with arbitrary  $\tau > 1$ .*

*Proof.* From Theorem 5.1 we obtain  $R_\alpha g^\delta \rightarrow \hat{g}^0$  for  $\delta \rightarrow 0$  and  $\|\widehat{R_\alpha g^\delta} - \hat{g}^0\|_1 = O\left(\delta^{\frac{2u}{u+1}}\right)$  if  $g^0 \in \mathcal{Y}^{1-u}$ . The  $\mathbb{L}^2$ -term in Equation (56) is actually much better behaved and can be estimated similarly. Now Equation (56) yields the result.  $\square$

Sharp images  $L(R_\alpha g^\delta)$  are recovered from noisy and blurred images  $g^\delta$  in two independent steps: First, the Fourier soft-shrinkage operator  $R_\alpha$  is applied. Then, the non-linear inversion operator  $L$  is applied. If the discrepancy principle is used, the parameter  $\alpha$  can be exactly computed with  $O(N \log N)$  operations, where  $N$  is the total number of pixels in the image, see [27]. The operator  $L$  is independent of  $\alpha$ . This two-step process [34], where data is first mollified and then inverted, is called a range-mollification [35] or pre-whitening method [3]. Figure 3 demonstrate some results. For a full discussion, we refer to [27].

## 8 Application to wavelet denoising

We briefly demonstrate the applicability of the theory developed in Section 6 to wavelet denoising. Our main aim is to show that the discrepancy principle does not only produce theoretically order-optimal results, but that the results actually look good, i.e. the constants in the estimates do not explode. We do not intend to compare our method with state-of-the-art techniques, which are, as we think, much more elaborate.

As basis of our test we pick the same image as in Section see Figure 3a). We have added 5%, 10%, 15% and 20% of normally distributed noise, where the percentage numbers specify the relative error  $\|g - g^\delta\|_2 / \|g\|_2$ . We apply the discrete 2d transform using Daubechis db10 filter coefficients and shrink the detail coefficients by the unique value  $\alpha$  for which

$$\|R_\alpha g^\delta - g^\delta\|_{\mathbb{L}^2(\Omega)} = \tau \delta$$

with  $\tau = 1.1$  and  $R_\alpha$  the wavelet shrinkage operator. Since the discrete wavelet transform defines an  $\mathbb{L}^2$ -isometry, the  $\mathbb{L}^2$ -norm can be efficiently computed in wavelet space. In fact, one can easily derive an  $O(N)$  algorithm for finding  $\alpha$ , where  $N$  is the total number of pixels in the image.

The results are presented in Figure 4. As expected, the denoised image converges to the original one for vanishing data error. The convergence rate results of Theorem 6.1 cannot be easily reproduced. In fact, the observed rates for  $\delta \rightarrow 0$  are  $O(\delta)$ , which is better than the theoretically optimal rate. What is the reason for this apparent contradiction? Being finite dimensional, the discretized version only *approximates* the underlying continuous shrinkage problem. Finite dimensional problems always exhibit  $O(\delta)$  behavior. Thus, the numerical results do not contradict the theory, but the numerical domain where the discrete problem approximates the continuous counterpart well enough is too small.

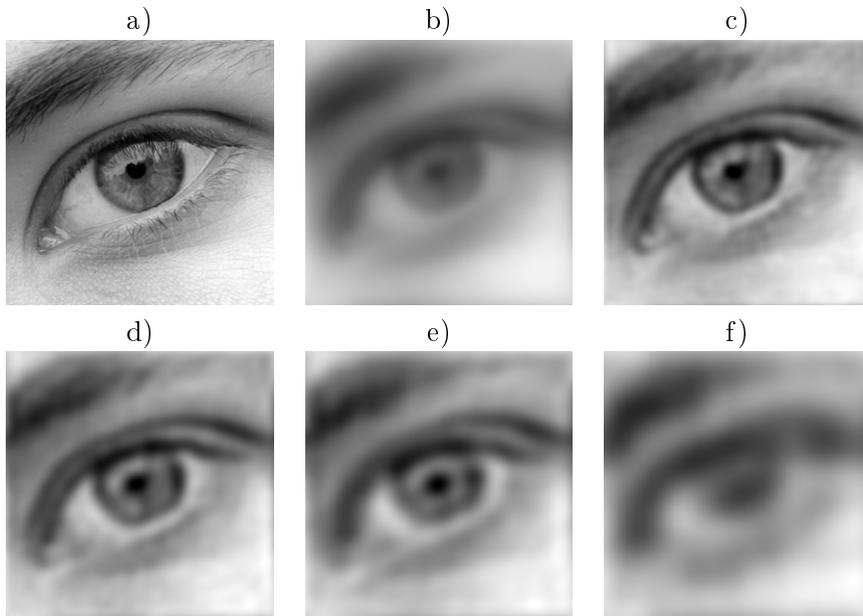


Figure 3: Blind deconvolution examples. a) original image, b) blurred image, c)-f) deconvolution results obtained by applying the two-step algorithm to the blurred image with additional 0.3%, 1.0%, 3% and 10% noise, respectively.

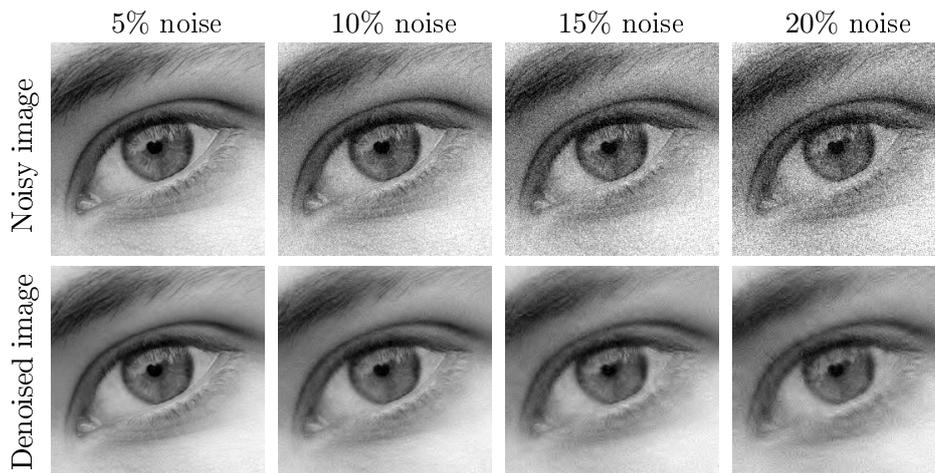


Figure 4: Wavelet denoising with four different noise levels. The images have been denoised by means of wavelet soft-shrinkage, where the shrinkage parameter has been computed from the discrepancy principle with  $\tau = 1.1$ .

## 9 Conclusions

We have investigated the problem of approximating a function  $\psi^0$  in the space  $\mathbb{L}^1(\mathbb{R}^d)$  given only noisy data  $\psi^\delta$  with  $\|\psi^0 - \psi^\delta\|_2 \leq \delta$ . The corresponding Tikhonov functional is minimized by soft shrinking the function  $\psi^\delta$  by some value  $\alpha$  and the minimizers converge to the true solution  $\psi^0$  for  $\delta \rightarrow 0$  for an appropriate choice of  $\alpha$ , e.g.  $\alpha \sim \delta$ . Thus, soft shrinkage is a convergent regularization method for the problem. As for all ill-posed problems, the convergence rate can be arbitrarily bad in general unless some source condition is satisfied. The natural source condition is  $\psi^0 \in \mathbb{L}^{1-u}(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$  for some  $0 < u < 1$ , leading to a convergence rate of  $O(\delta^{\frac{2u}{1+u}})$  in the  $\mathbb{L}^1$ -norm. This rate is in fact order-optimal. The discrepancy principle as a posteriori parameter choice rule is of special importance for practical applications since it does not require the knowledge of the exact  $u$  for which the source condition is satisfied. Nevertheless, the discrepancy principle implies the same convergence rates as a priori rules, in particular, it is of optimal order.

All results for this rather abstract minimization problem can be carried over to practically relevant problems by combining the shrinkage operator by a linear isometry. We have demonstrated this for the continuous Fourier transform with application to blind deconvolution and for the wavelet transform with application to wavelet denoising. In case of the Fourier transform, the corresponding source spaces  $\mathcal{Y}^t$  contain functions with Fourier transform in  $\mathbb{L}^t(\mathbb{R}^d)$ ,  $0 < t < 1$ , and do not imply any smoothness in terms of differentiability. In case of the wavelet transform, the source spaces  $\mathcal{Y}^p$  are Besov spaces  $B_{pp}^s(\Omega)$  along the line  $s = \frac{2-p}{p}$  with  $0 < p < 1$ . Images are reconstructed in  $B_{11}^1(\Omega)$ , a space which describes images well since it is close to the space of bounded variation. We stress that order optimality, convergence rates and the applicability of the discrepancy principle carry over to the applications as well.

In principle, the general approach to data smoothing is applicable to a wide range of even non-linear ill-posed problems  $A(f) = g$ . We have demonstrated this for blind deconvolution. In a two-step process the data is first smoothed into the range of the operator and then an appropriate (non-linear) inversion operator is applied. This procedure has the advantage of decoupling regularization (data smoothing) and inversion. Thus, it is rather fast; for instance, each blind deconvolution reconstruction takes only the fraction of a second. Our future goal is to exploit these advantages for a greater range of inverse problems.

## References

- [1] F. Natterer. *The Mathematics of Computerized Tomography*. B.G. Teubner, Stuttgart, 1986.
- [2] Ronny Ramlau and Gerd Teschke. Regularization of sobolev embedding operators and applications to medical imaging and meteorological data. Part I: Regularization of sobolev embedding operators. *Sampling Theory in Signal and Image Processing*, 3(2):175–196, May 2004.

- [3] P. Jonas and A.K. Louis. A sobolev space analysis of linear regularization methods for ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, 9(1):59–74, 2001.
- [4] Albert Cohen, Marc Hoffmann, and Markus Reiss. Adaptive wavelet galerkin methods for linear inverse problems. *SIAM J. Numerical Analysis*, 42, 2004.
- [5] Esther Klann, Peter Maass, and Ronny Ramlau. Two-step regularization methods for linear inverse problems. *Journal of Inverse and Ill-Posed Problems*, 14(6):583–607(25), 2006.
- [6] Esther Klann and Ronny Ramlau. Regularization by fractional filter methods and data smoothing. *submitted*, 2007.
- [7] Ronny Ramlau and Gerd Teschke. Regularization of sobolev embedding operators and applications to medical imaging and meteorological data. Part II: Regularization incorporating noise with applications in medical imaging and meteorological data. *Sampling Theory in Signal and Image Processing*, 3(3):205–226, September 2004.
- [8] Leonid I. Rudin, Stanley J. Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [9] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.
- [10] David C. Dobson and Curtis R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM Journal on Numerical Analysis*, 34:1779–1971, 1997.
- [11] G. Chavent and K. Kunisch. Regularization of linear least squares problems by total bounded variation. *ESAIM, Control, Optimization and Calculus of Variations*, 2:359–376, 1997.
- [12] Martin Burger and Stanley Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20:1411–1421, 2004.
- [13] D. Dobson and O. Scherzer. Analysis of regularized total variation penalty methods for denoising. *Inverse Problems*, 12:601–617, 1996.
- [14] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Math. Imaging and Vision*, 20:89 – 97, 2004.
- [15] Ronald A. DeVore and Bradley J. Lucier. Fast wavelet techniques for near-optimal image processing. In *Communications - Fusing Command, Control and Intelligence*, volume 3 of *Military Communications Conference, MIL-COM '92*, pages 129–135. IEEE, 1992.

- [16] Antonin Chambolle, Ronald A. DeVore, Namyong Lee, and Bradley J. Lucier. Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7:319–335, 1998.
- [17] Antonin Chambolle and Bradley J. Lucier. Interpreting translation-invariant wavelet shrinkage as a new image smoothing scale space. *IEEE Transactions on Image Processing*, 10:993–1000, 2001.
- [18] Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald A. DeVore. Harmonic analysis of the space  $BV$ . *Revista Matemática Iberoamericana*, 19(1):235–263, 2003.
- [19] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space  $BV(\mathbb{R}^2)$ . *Am. J. Math.*, 121:587–628, 1999.
- [20] I. daubechies and G. Teschke. Variational image restauration by means of wavelets: simultaneous decomposition,deblurring and denoising. *Appl. Comput. Harmonic. Anal.*, 19(1):1–16, 2005.
- [21] David L. Donoho. Denoising via soft thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.
- [22] David L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Annals of Statistics*, 24(2), 1996.
- [23] David L. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. Technical Report 402, Dept. of Statistics, Stanford University, 1992.
- [24] Kristian Bredies, Dirk A. Lorenz, and Peter Maass. Mathematical concepts of multiscale smoothing. *Applied and Computational Harmonic Analysis*, 19(2):141–161, August 2005.
- [25] Albert Cohen, Ronald DeVore, Gerard Kerkyacharian, and Dominique Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(2):167–191, 2001.
- [26] R.E. Showalter. *Monotone operators in Banach spaces and nonlinear partial differential equations*, volume 49 of *Math. Surveys and Monographs*. Amer. Math. Soc., 1997.
- [27] Lutz Justen. *Blind Deconvolution: Theory, Regularization and Applications*. PhD thesis, University of Bremen, 2006.
- [28] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [29] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, USA, 1999.

- [30] Alfred Karl Louis, Peter Maaß, and Andreas Rieder. *Wavelets: Theory and Application*. Wiley, Chichester, 1997.
- [31] Albert Cohen. *Numerical Analysis of Wavelet Methods*. Elsevier Science B.V., 2003.
- [32] Martin Burger and Otmar Scherzer. Regularization methods for blind deconvolution and blind source separation problems. *Mathematics of Control, Signals and Systems*, 14:358–383, 2001.
- [33] Lutz Justen and Ronny Ramlau. A non-iterative regularization approach to blind deconvolution. *Inverse Problems*, 22(3):771–800, 2006.
- [34] Esther Klann. *Regularization of linear ill-posed problems in two steps: combination of data smoothing and reconstruction methods*. PhD thesis, University of Bremen, 2005.
- [35] M. Hegland and R.S. Anderssen. A mollification framework for improperly posed problems. *Numerische Mathematik*, 78(4):549–575, February 1998.