Inverse Problems

PAPER

Multiscale hierarchical decomposition methods for ill-posed problems

To cite this article: Stefan Kindermann *et al* 2023 *Inverse Problems* **39** 125013

IOP Publishing

Inverse Problems

Inverse Problems **39** (2023) 125013 (36pp)

https://doi.org/10.1088/1361-6420/ad05de

Multiscale hierarchical decomposition methods for ill-posed problems

Stefan Kindermann[1], **Elena Resmerita**[2] **and Tobias Wolf**[2,*]

[1] Industrial Mathematics Institute, Johannes Kepler University Linz, Linz, Austria
[2] Institute of Mathematics, University of Klagenfurt, Klagenfurt, Austria

E-mail: tobias.wolf@aau.at

Received 18 April 2023; revised 25 September 2023
Accepted for publication 23 October 2023
Published 7 November 2023

Abstract

The multiscale hierarchical decomposition method (MHDM) was introduced in Tadmor *et al* (2004 *Multiscale Model. Simul.* **2** 554–79; 2008 *Commun. Math. Sci.* **6** 281–307) as an iterative method for total variation (TV) regularization, with the aim of recovering details at various scales from images corrupted by additive or multiplicative noise. Given its success beyond image restoration, we extend the MHDM iterates in order to solve larger classes of linear ill-posed problems in Banach spaces. Thus, we define the MHDM for more general convex or even nonconvex penalties, and provide convergence results for the data fidelity term. We also propose a flexible version of the method using adaptive convex functionals for regularization, and show an interesting multiscale decomposition of the data. This decomposition result is highlighted for the Bregman iteration method that can be expressed as an adaptive MHDM. Furthermore, we state necessary and sufficient conditions when the MHDM iteration agrees with the variational Tikhonov regularization, which is the case, for instance, for one-dimensional TV denoising. Finally, we investigate several particular instances and perform numerical experiments that point out the robust behavior of the MHDM.

Keywords: ill-posed problem, multiscale regularization, total variation

1. Introduction

In their influential works from 2004 and 2008, Tadmor *et al* [1, 2] introduced a multiscale decomposition method for image denoising, deblurring and segmentation, based on the popular

* Author to whom any correspondence should be addressed.

total variation (TV) model of Rudin, Osher and Fatemi (ROF) [3]. Recall that ROF decomposes an image $f \in L^2(\Omega)$ in cartoon and texture as $f = u_\lambda + v_\lambda$ such that

$$(u_\lambda, v_\lambda) = \arg\min_{u+v=f} \left\{ \lambda \|v\|_{L^2}^2 + |u|_{\mathrm{TV}} \right\}, \tag{1.1}$$

with $v_\lambda \in L^2(\Omega)$ and $u_\lambda \in BV(\Omega) = \{u \in L^2(\Omega) : |u|_{\mathrm{TV}} < \infty\}$, where $|\cdot|_{\mathrm{TV}}$ denotes the TV seminorm given by $|u|_{\mathrm{TV}} := \sup\{\int_\Omega u \ \mathrm{div}\varphi : \varphi \in C_0^\infty(\Omega) \text{ and } \|\varphi\|_\infty \leqslant 1\}$. Here Ω is a bounded and open set in $\mathbb{R}^2$. While the main features in natural images are very well restored via ROF, the texture at various scales might not be optimally recovered. The multiscale hierarchical decomposition method (MHDM) copes with this difficulty by decomposing an image into a sum of multiple images, each of these containing features of the original image at a different scale. Thus, besides extracting a cartoon representation of the original image, it allows recovering more oscillatory image components. One of the main advantages is the relatively simple modeling involved in the procedure. Instead of employing more complicated and numerically expensive penalty terms, the improvement is achieved by 'zooming-in': what is considered noise and texture at one scale, can be regarded as cartoon at a finer scale. The explicit decomposition of images into parts that contain increasingly more subtle features has made the method attractive to solve plenty of other problems. Examples can be found in the fields of nonlinear partial differential equations [4], image registration [5, 6], graph theory [7], compressed sensing, deconvolution of the Helmholtz filter and linear regression—see the PhD thesis [8]. Note that the latter concerns actually a general MHDM, but for solving linear inverse problems in finite dimension. Moreover, the approach in [5] corresponds to a wider range of possible applications apart from image registration since it employs nonlinear operators, not necessarily quadratic data fidelities, and powers of seminorms as penalties.

In the sequel, we formulate the MHDM in Banach spaces and present the state of the art to the best of our knowledge.

Let X be a Banach space and $J : X \to [0, \infty]$ be a proper and lower semicontinuous functional which is bounded from below. Let $T \in \mathcal{L}(X, H)$ be an ill-posed linear operator with values in a Hilbert space H, and fix $f \in H$. Generally, we are interested in solving

$$\begin{cases} \text{minimize} & J(x) \\ s.t. & Tx = f. \end{cases} \tag{1.2}$$

We will furthermore assume that problem (1.2) is non-degenerate, that is, there exists an element $x^\dagger \in \mathrm{dom}\, J := \{x \in X : J(x) < \infty\}$ such that

$$Tx^\dagger = f. \tag{1.3}$$

Throughout this work, we assume that the generalized Tikhonov functional

$$F_\lambda(u) := \frac{\lambda}{2} \|Tu - y\|^2 + J(u) \tag{1.4}$$

admits a minimizer for all $\lambda > 0$ and $y \in H$. Standard conditions when this holds true can be found, e.g. in [9]. Moreover, we denote a minimizer of the generalized Tikhonov functional as x_λ, i.e.

$$x_\lambda := \arg\min_{u \in X} F_\lambda(u) = \arg\min_{u \in X} \frac{\lambda}{2} \|Tu - y\|^2 + J(u). \tag{1.5}$$

We approach (1.2) by the MHDM that works as follows: choose a sequence $(\lambda_n)_{n\in\mathbb{N}_0}$ of positive real numbers and compute

$$u_0 \in \arg\min_{u\in X} \frac{\lambda_0}{2}\|Tu-f\|^2 + J(u). \tag{1.6}$$

Denote the residual $v_0 = f - Tu_0$ and set $x_0 = u_0$. Next, compute iteratively for $n = 1, \ldots$

$$\begin{aligned} &u_n \in \arg\min_{u\in X} \frac{\lambda_n}{2}\|v_{n-1} - Tu\|^2 + J(u), \\ &x_n = x_{n-1} + u_n \qquad \text{and} \\ &v_n = v_{n-1} - Tu_n = f - Tx_n. \end{aligned} \tag{1.7}$$

The resulting sequence $(x_n)_{n\in\mathbb{N}_0}$ with $x_n = \sum_{i=1}^{n} u_i$ is considered as an approximation of $x^\dagger$, thus yielding a scale decomposition depending on the choice of J and λ_n.

Note that (1.7) can be rewritten as

$$x_n \in \arg\min_{x\in X} \frac{\lambda_n}{2}\|Tx-f\|^2 + J(x - x_{n-1}). \tag{1.8}$$

If X is a Hilbert space and $J = \frac{\|\cdot\|^2}{2}$, this procedure reads

$$x_n \in \arg\min_{x\in X} \lambda_n\|Tx-f\|^2 + \|x - x_{n-1}\|^2, \tag{1.9}$$

which is the nonstationary Tikhonov regularization—see [10, 11] for detailed convergence and error estimate results, as well as [12] concerning the inverse scale space method as an asymptotic formulation of the method. In fact, one recognizes in (1.9) the particular quadratic setting for another prominent approach of inverse problems, namely the nonstationary augmented Lagrangian method [13] known also as the Bregman iteration [14]. Note, however, that these methods usually differ from the MHDM method for non-quadratic penalties J. Before describing the latter, let us recall some definitions from convex analysis. If the functional J is convex, one defines the subgradient of J at a point $x_0 \in \operatorname{dom} J$ as

$$\partial J(x_0) = \{x^* \in X^* : \langle x^*, x - x_0\rangle \leqslant J(x) - J(x_0) \text{ for all } x \in X\}, \tag{1.10}$$

where X^* stands for the dual space of X. Furthermore, for any point $x_1 \in \operatorname{dom} J$, the Bregman distance of x_0 and x_1 with respect to $x^* \in \partial J(x_0)$ is denoted by

$$D_J^{x^*}(x_1, x_0) = J(x_1) - J(x_0) - \langle x^*, x_1 - x_0\rangle. \tag{1.11}$$

Now we are in a position to recall the Bregman iteration: For some sequence $(\lambda_n) \subset (0,\infty)$ and for any $n \in \mathbb{N}$, let

$$x_n \in \arg\min_{x\in X} \frac{\lambda_n}{2}\|Tx-f\|^2 + D_J^{p_{n-1}}(x, x_{n-1}), \tag{1.12}$$

with $x_0 = 0$ and $p_0 = 0$, where in each step one chooses

$$p_n = \lambda_n T^*(f - Tx_n) + p_{n-1} \in \partial J(x_n).$$

Here $T^* : H \to X^*$ stands for the adjoint of the operator T. While for the related method (1.12) comprehensive convergence results exist in the corresponding references mentioned above, the situation is different for the MHDM defined with a non-quadratic penalty J in infinite-dimensional spaces. Intriguingly, there is no convergence result for the sequence of iterates (x_n) apart from the denoising case (when T is the identity), which is a consequence of the weak-/strong convergence for the residual $(Tx_n - f)$ (see [1, 2, 5]) or of the residual error estimates in [15]. As regards the penalty J, this was assumed so far to be a (power of a) seminorm.

Another interesting open question was raised in [1], whether the MHDM iterate x_n coincides with the solution x_{λ_n} of Tikhonov regularization (1.5) corresponding to the parameter λ_n.

In general, this is not the case. However, there are situations when the answer is positive, as can be seen in section 4.

Thus, the main contributions of this study are as follows. First of all, we extend the existing convergence results regarding the residual to the case when the penalty is a more general convex function than a (power of a) seminorm or when it belongs to a class of nonconvex functions. Furthermore, we propose a generalization of MHDM by empowering the penalty to be adaptive, and point out a couple of specific penalties that yield well known methods for solving (1.2). For instance, as a side result that is interesting in itself, we formulate the Bregman iteration (1.12) as a generalized MHDM with appropriate adaptive penalties, and obtain a curious multiscale norm decomposition of the data in terms of (Tu_n) and symmetric Bregman distances $D_J^{\mathrm{sym}}(x_n, x_{n-1})$. Since the topic of generalized MHDM with new, meaningful, adaptive penalties is quite challenging, we will consider it for future research. Moreover, we state necessary and sufficient conditions for the MHDM to agree with the Tikhonov regularization. We verify these conditions for one-dimensional TV-denoising, as well as for particular situations in two-dimensional TV-denoising and for finite-dimensional ℓ^1-regularization. In particular, we emphasize that the so-called positive cone condition [16] for the operator T in the ℓ^1-regularization case does allow one to compare the MHDM iterate x_n to the solution x_{λ_n} of Tikhonov regularization (1.5) corresponding to the regularization parameter λ_n used in the MHDM. Last but not least, we focus on several examples to understand how MHDM performs theoretically and computationally. The numerical experiments show a robust behavior of the MHDM with respect to the involved parameters. Moreover, they give us hope that, under suitable assumptions which need to be found, the Multiscale Hierarchical Decomposition Method does converge in a more general framework, that is when $T \neq Id$, where Id denotes the identity operator.

The structure of this work is as follows. Section 2 presents the convergence of the residual in the case of some convex or even nonconvex penalties. Thereafter, we suggest in section 3 a generalization of the iterative method and derive a decomposition result of the data. The comparison of the MHDM and the generalized Tikhonov regularization can be found in section 4, while section 5 presents the numerical experiments.

2. Convergence of the residual

As mentioned in the introduction, no general result regarding the convergence of the MHDM iterates has been shown when the problem $Tx = f$ is ill-posed. However, if the problem is well-posed, that is, T is a bijective linear operator with continuous inverse, convergence of the iterates (x_n) is a consequence of the convergence of the residual. For the setting when J is a power of a seminorm, it was shown in [5] that the (not necessarily quadratic) residual converges. Nonetheless, by adding the term $J(x)$ to (1.8), in [5] it is proved that the iterates (x_n) of the resulting *tight MHDM* (cf (3.6) below) converge on subsequences to a solution of (1.2). Similar results were derived in [15] for the iterates of a refined version of the tight MHDM that uses two different penalties, namely J for the components u_n and R for the sum

x_n of the components u_n. Note that [15] established also error estimates for the residual of the MHDM and of its tighter versions.

In this section, we extend the latter result under the assumption of a generalized triangle inequality on the penalty J. This will lead to convergence rates of the residual for large classes of penalty functions, including certain nonconvex functions. Additionally, we establish a weak convergence result for the residual in a complementary case of general convex functions (which do not necessarily satisfy a generalized triangle inequality).

2.1. The case of exact data

We start by showing the results under the assumption of exact data f.

Theorem 2.1. *Let $(x_n)_{n\in\mathbb{N}_0}$ be the sequence generated by (1.6)–(1.7), and let (1.3) hold.*

(i) Assume that J is minimal at 0 and that there is $C \geqslant 1$ such that

$$J(x-y) \leqslant C(J(x)+J(y)) \tag{2.1}$$

for all $x, y \in X$. If λ_n is chosen such that $2C\lambda_{n-1} \leqslant \lambda_n$ for all $n \in \mathbb{N}$, then the residual satisfies

$$\|f - Tx_n\| \leqslant \left(4C\frac{J(x^\dagger)}{\lambda_0 (n+1)}\right)^{\frac{1}{2}} \tag{2.2}$$

for all $n \in \mathbb{N}_0$.

(ii) Assume J is convex and dom J is dense in X. Moreover, let one of the following conditions hold.

(a) If J is minimal at 0, then the residuals are monotonically decreasing.

(b) If $J(0) < \infty$ and λ_n is chosen such that

$$\sum_{n=0}^{\infty} \frac{1}{\lambda_n} < \infty, \tag{2.3}$$

then the residuals satisfy

$$\|f - Tx_n\|^2 \leqslant 2\tilde{C}\sum_{j=0}^{n}\frac{1}{\lambda_n},$$

where $\tilde{C} = J(0) - \inf_{x\in X} J(x)$.

In both cases (a) and (b), $((f - Tx_n))_{n\in\mathbb{N}_0}$ is bounded, and every weak limit point is in the kernel of T^*. In particular, this means that $T^*(f - Tx_n)$ converges to 0 in the weak--topology of X^*.*

Proof. (i) Let $n \in \mathbb{N}$. By the optimality of u_n defined in (1.6) and (1.7) by comparing to $u = 0$ it holds

$$\frac{\lambda_n}{2}\|f - Tx_n\|^2 + J(u_n) \leqslant \frac{\lambda_n}{2}\|f - Tx_{n-1}\|^2 + J(0). \tag{2.4}$$

On the other hand, comparing to $u = x^\dagger - x_{n-1}$, we obtain

$$\frac{\lambda_n}{2}\|f - Tx_n\|^2 + J(u_n) \leqslant J\left(x^\dagger - x_{n-1}\right). \tag{2.5}$$

In particular, the minimality of J at 0 and (2.4) imply that $\|f - Tx_n\|$ is decreasing. Using (2.1), it holds for $k \in \mathbb{N}$,

$$\begin{aligned}
\frac{\lambda_k}{2}\|f - Tx_k\|^2 + J\left(x^\dagger - x_k\right) &= \frac{\lambda_k}{2}\|f - Tx_k\|^2 + J\left(x^\dagger - x_{k-1} - u_k\right) \\
&\leqslant \frac{\lambda_k}{2}\|f - Tx_k\|^2 + C\left(J\left(x^\dagger - x_{k-1}\right) + J(u_k)\right) \\
&= \frac{\lambda_k}{2}\|f - Tx_k\|^2 + J(u_k) + (C-1)J(u_k) + CJ\left(x^\dagger - x_{k-1}\right) \\
&\overset{(2.5)}{\leqslant} (C-1)J(u_k) + (C+1)J\left(x^\dagger - x_{k-1}\right) \\
&\leqslant 2CJ\left(x^\dagger - x_{k-1}\right),
\end{aligned}$$

where the last inequality follows from (2.5) and $\frac{\lambda_k}{2}\|f - Tx_k\|^2 \geqslant 0$. By using this and the choice of λ_k, we can conclude

$$\frac{1}{2}\|f - Tx_k\|^2 + \frac{J\left(x^\dagger - x_k\right)}{\lambda_k} \leqslant \frac{2C}{\lambda_k}J\left(x^\dagger - x_{k-1}\right) \leqslant \frac{1}{\lambda_{k-1}}J\left(x^\dagger - x_{k-1}\right). \tag{2.6}$$

Thus, we can repeatedly use (2.6) for $k = 0, \ldots, n$, sum up and obtain

$$(n+1)\frac{\|f - Tx_n\|^2}{2} + \frac{J\left(x^\dagger - x_n\right)}{\lambda_n} \leqslant \sum_{k=0}^{n}\left(\frac{\|f - Tx_k\|^2}{2}\right) + \frac{J\left(x^\dagger - x_n\right)}{\lambda_n} \leqslant \frac{2C}{\lambda_0}J\left(x^\dagger\right),$$

where the first inequality is implied by the monotonicity of the residual.
This yields (2.2).

(ii) Let $n \in \mathbb{N}$. Then by the optimality of u_n, we obtain

$$\frac{\lambda_n}{2}\|f - Tx_n\|^2 + J(u_n) \leqslant \frac{\lambda_n}{2}\|f - Tx_{n-1}\|^2 + J(0). \tag{2.7}$$

This means $\|f - Tx_n\|^2 \leqslant \|f - Tx_{n-1}\|^2 + 2\frac{J(0)-J(u_n)}{\lambda_n}$. If now $J(0)$ is minimal, this means that $\|f - Tx_n\|$ is decreasing. Otherwise, we inductively arrive at

$$\|f - Tx_n\|^2 \leqslant \|f\|^2 + 2\sum_{k=0}^{n}\frac{J(0) - J(u_k)}{\lambda_k} \leqslant \|f\|^2 + 2\sum_{k=0}^{n}\frac{J(0)}{\lambda_k}.$$

By (2.3), this implies that in both cases the sequence $(f - Tx_n)$ is bounded. Therefore, it admits a weakly convergent subsequence $(f - Tx_{n_k})$. Let v be its limit. Note that by the optimality condition of u_n and due to the convexity of J, it holds

$$\lambda_n T^*(f - Tx_n) \in \partial J(u_n).$$

This means

$$\begin{aligned}\langle T^{*}(f-Tx_n),z\rangle_X &\leqslant \frac{J(z)-J(u_n)}{\lambda_n}+\langle f-Tx_n,Tu_n\rangle_H\\ &\leqslant \frac{J(z)}{\lambda_n}+\|f-Tx_n\|\,\|T(x_n-x_{n-1})\|\\ &\leqslant \frac{J(z)}{\lambda_n}+K,\end{aligned} \tag{2.8}$$

for some $K\geqslant 0$, because $\|f-Tx_n\|$ and therefore $\|Tx_n\|$ are bounded. Considering now (2.8) with x_n replaced by the subsequence x_{n_k} and letting $k\to\infty$ yield

$$\langle T^{*}v,z\rangle \leqslant K \tag{2.9}$$

for all z with $J(z)<\infty$. Now assume $T^{*}v\neq 0$. Then, for all $n\in\mathbb{N}$, there is z_n with $\langle T^{*}v,z_n\rangle\geqslant n$. Fix $\varepsilon>0$. By the density of dom J which is equivalent to the density in the weak topology by Mazur's Lemma [17, p 6] (as dom J is convex), we can find z_n^{ε} with $J(z_n^{\varepsilon})<\infty$ and $|\langle T^{*}v,z_n-z_n^{\varepsilon}\rangle|\leqslant\varepsilon$. Thus $\langle T^{*}v,z_n^{\varepsilon}\rangle\geqslant n-\varepsilon$ holds, contradicting (2.9). Since this reasoning can be applied to any subsequence, the claim follows.

□

Remark.

1. If $J(x)=|x|_{se}^{p}$ for a seminorm $|\cdot|_{se}$ and $p\in(0,\infty)$, then condition (2.1) holds for $C=2^{\max\{0,p-1\}}$. Furthermore, if $(p_n)_{n\in\mathbb{N}}$ is a bounded sequence in $(0,\infty)$, then $J(x)=\sum_{n=1}^{\infty}|x_n|^{p_n}$ satisfies (2.1) with $C=2^{\max\{0,\sup p_n-1\}}$. The reader is referred to [18] for promoting sparsity by employing the latter functional J.
2. In the situation of part (*i*) in the previous theorem, strong convergence of the residual is also obtained if $(\lambda_n)_{n\in\mathbb{N}_0}$ is an arbitrary sequence increasing to ∞ (a proof is given for a more general version of the algorithm in lemma 3.2). However, we do not obtain convergence rates without additionally assuming the rate of increase to be at least geometric.
3. Estimate (2.2) also holds for general distance functions: Let $d:H\times H\to[0,\infty)$ be a function such that $d(x,x)=0$ for all $x\in X$. If we replace the Hilbert space norm in (1.6) and (1.7) by d, i.e. we consider the iteration

$$u_n\in\operatorname*{arg\,min}_{u\in X}\lambda_n d(Tu,v_{n-1})+J(u),$$

then, under the same assumptions as in part (*i*) of 2.1, the estimate

$$d(Tx_n,f)-\inf_{x\in X}d(Tx,f)\leqslant 2C\frac{J(x^{\dagger})}{\lambda_0(n+1)} \tag{2.10}$$

holds. This means, we have found convergence rates of the residual for all penalties considered in section 2 of [5]. Furthermore, by making minor adaptions in the notation of the proof, (2.10) also holds for nonlinear operators T. However, for nonconvex data fidelity terms or nonlinear operators, computing global minimizers of the corresponding Tikhonov functionals might be impossible or too expensive for practical applications. In those cases, one might consider alternative ways to regularize the ill-posed problem.

2.2. The noisy data case

Assume now that instead of exact data f, we are given a noisy measurement f^{δ}. Typically, regularization methods behave semi-convergent in this situation. This means that the true solution is approached initially, but then the distance between the iterates and the true solution eventually increases. To overcome this issue, the iteration is terminated early according to some meaningful rule. We will use the discrepancy principle as a stopping criterion and show convergence of the residual as the noise level δ approaches 0.

Let us assume the case of additive Gaussian noise, i.e.

$$\left\|f^{\delta}-f\right\| \leqslant \delta \tag{2.11}$$

with some $\delta > 0$. For $n \in \mathbb{N}$, denote by $x_n^{\delta} = \sum_{j=0}^{n} u_j^{\delta}$ the iterate of the MHDM with data f^{δ} instead of f.

Lemma 2.2. *Assume that f and f^{δ} verify (2.11). If J satisfies (2.1) and the sequence $(\lambda_n)_{n\in\mathbb{N}_0}$ is chosen such that $2C\lambda_{n-1} \leqslant \lambda_n$, then the following holds, for all $n \in \mathbb{N}$:*

$$\left\|Tx_n^{\delta}-f^{\delta}\right\|^2 \leqslant 4C\frac{J\left(x^{\dagger}\right)}{\lambda_0\left(n+1\right)} + \delta^2. \tag{2.12}$$

Proof. Let $n \in \mathbb{N}$. By the optimality of u_n^{δ} for (1.7) with f replaced by f^{δ}, it is

$$\frac{\lambda_n}{2}\left\|f^{\delta}-Tx_n^{\delta}\right\|^2 + J\left(u_n^{\delta}\right) \leqslant \frac{\lambda_n}{2}\left\|f^{\delta}-f\right\|^2 + J\left(x^{\dagger}-x_{n-1}^{\delta}\right) \leqslant \frac{\lambda_n}{2}\delta^2 + J\left(x^{\dagger}-x_{n-1}^{\delta}\right).$$

By using analogous reasoning as in the proof of theorem 2.1 (i), the claim follows. □

We consider the following discrepancy principle: Choose some $\tau > 1$ and let the index

$$n^*(\delta) = \max\left\{n \in \mathbb{N} : \left\|Tx_n^{\delta}-f^{\delta}\right\|^2 \geqslant \tau\delta^2\right\}. \tag{2.13}$$

Note that $n^*(\delta)$ is well-defined by (2.12).

Theorem 2.3. *If under the assumption of lemma 2.2, the iteration is stopped at index $n^*(\delta)+1$, then $Tx_{n^*(\delta)+1}^{\delta} \to f$ as $\delta \to 0$.*

Proof. By (2.13), it is

$$\left\|Tx_{n^*(\delta)+1}^{\delta}-f^{\delta}\right\| \leqslant \tau^{\frac{1}{2}}\delta.$$

Therefore, the estimate

$$\left\|Tx_{n^*(\delta)+1}^{\delta}-f\right\| \leqslant \left\|Tx_{n^*(\delta)+1}^{\delta}-f^{\delta}\right\| + \left\|f-f^{\delta}\right\| \leqslant \tau^{\frac{1}{2}}\delta + \delta$$

yields the result. □

Remark. In the situation of part (*ii*) in theorem 2.1, we also get that $T^*(f^{\delta}-Tx_n^{\delta})$ converges to 0 in the weak-*-topology. Yet, it is not clear how to define a meaningful stopping index, as there are no convergence rates available.

2.3. Maximum entropy regularization

In this section, we give an example where the conditions of theorem 2.1 are not satisfied and the residual of the MHDM does not converge.

Let $\Omega \subset \mathbb{R}^n$ be bounded. The negative Boltzman–Shannon entropy is defined as $S : L^1(\Omega) \to \mathbb{R} \cup \{\infty\}$,

$$S(x) = \begin{cases} \int_\Omega x(t) \log x(t) \, \mathrm{d}t & \text{if } x \geqslant 0 \text{ a.e. and } x \log x \in L^1(\Omega) \\ \infty & \text{else.} \end{cases} \tag{2.14}$$

Here we use the convention $0 \log 0 = 0$. This functional can be employed in regularization methods to enforce non-negativity of approximate solutions—see, e.g. [19–21], as well as the survey [22]. For an operator $T \in \mathcal{L}(X, H)$, the variational regularization (1.5) with penalty $J = S$ is well-defined due to the coercivity and lower semicontinuity of the entropy with respect to the weak topology of $L^1(\Omega)$. This functional is of particular interest in our context, because it does not satisfy the conditions of theorem 2.1.

Lemma 2.4. *Let $X = L^1(\Omega)$ and let $J = S$. Then J does not satisfy (2.1), for any $C \geqslant 1$.*

Proof. Let $x(t) = y(t) = \mathrm{e}^{-1}$ for all $t \in \Omega$. Since $x, y \in \operatorname{dom} S$ and $S(x - y) = 0$, one has

$$S(x) + S(y) = -2 \int_\Omega \mathrm{e}^{-1} \mathrm{d}t < 0.$$

Thus, (2.1) can not hold for any $C \geqslant 1$. □

Since the domain of S is strictly contained in $L^1_+(\Omega) := \{u \in L^1 : u \geqslant 0 \text{ a.e.}\}$, the density assumption in theorem 2.1 does not hold either. Thus, the theorem can not be applied to the MHDM defined with the entropy penalty. In fact, one can show that the residual obtained in this setting does not necessarily converge. In order to see this, e.g. for the simple example of the identity operator, we recall the proximal mapping of the negative entropy. Using table 2 in [23] and proposition 12.22 in [24] pointwise, one has

$$prox_{\frac{S}{\lambda}}(y) = \underset{x \in X}{\arg\min} \frac{\lambda}{2} \|x - y\|^2 + S(x) = \frac{1}{\lambda} W(\lambda \exp(\lambda y - 1)) \tag{2.15}$$

where W denotes the principal branch of the Lambert W function.

Lemma 2.5. *Let $X = H = L^2(\Omega)$ and $T = Id$. Let furthermore $(\lambda_n)_{n \in \mathbb{N}_0}$ be a sequence of increasing positive numbers, and assume that $x^\dagger$ satisfies $0 \leqslant x^\dagger < \frac{1}{e}$ on a set E with positive measure. Then the MHDM iterates with S (restricted to L^2) as penalty term will be bounded away from $x^\dagger$ in the following sense:*

$$x_n(t) > \cdots > x_0(t) > x^\dagger(t), \quad \textit{for all } t \in E \textit{ and all } n \in \mathbb{N}.$$

Proof. Since $T = Id$, it is $u_n = \operatorname{prox}_{\frac{S}{\lambda_n}}(x^\dagger - x_{n-1})$ for $n \in \mathbb{N}_0$ and $x_{-1} = 0$. We can therefore use a one-dimensional calculation. Let $0 \leqslant x < \frac{1}{e}$ and $\lambda \geqslant 0$. Multiplying the inequality

$$\lambda e^{-1} > \lambda x$$

with $\exp(\lambda x)$ and applying W (which is strictly increasing and positive on $[0, \infty)$) yield

$$W(\lambda \exp(\lambda x - 1)) > W(\lambda x \exp(\lambda x)) = \lambda x,$$

or equivalently

$$\frac{1}{\lambda}W(\lambda\exp(\lambda x-1))>x.$$

Thus, because $x^{\dagger}(t)<e^{-1}$ on E, the previous inequality with $x^{\dagger}(t)$ instead of x and equation (2.15) imply

$$x_0(t)=u_0(t)=\left(\operatorname{prox}_{\frac{S}{\lambda_0}}(x^{\dagger})\right)(t)>x^{\dagger}(t),$$

independently of the choice of λ_0, for all $t\in E$. Since by (2.15) the increment

$$x_1-x_0=u_1=\operatorname{prox}_{\frac{S}{\lambda_1}}\left(x^{\dagger}-x_0\right)$$

is positive, we must have $x_1(t)>x_0(t)>x^{\dagger}(t)$ for all $t\in E$. The statement now follows by induction. □

The previous lemma shows that we cannot expect L^2-convergence of the residual (which in this case is the same as convergence of (x_n)) for the MHDM with entropy penalty, if $x^{\dagger}<e^{-1}$ on a set of positive measure. In fact, the residual does not converge even if the ground truth is bounded away from $\frac{1}{e}$. Indeed, assume (x_n) does converge to $x^{\dagger}$. If some iterate x_n satisfies $x_n>x^{\dagger}$ on a set of positive measure, then (2.15) implies that $x_{n+1}>x_n$ on that set, meaning that x_n cannot converge to $x^{\dagger}$. Otherwise, if $x_n\leqslant x^{\dagger}$ a.e. for all $n\in\mathbb{N}_0$, then convergence would yield $0\leqslant x^{\dagger}-x_{n_0}\leqslant e^{-1}$ on a set E on positive measure for some $n_0\in\mathbb{N}_0$. Using the same reasoning as in the proof of lemma 2.5 but with $x^{\dagger}-x_{n_0}$ instead of $x^{\dagger}$, one would obtain $u_{n_0+1}>x^{\dagger}-x_{n_0}$ on E. But this would be equivalent to $x_{n_0+1}>x^{\dagger}$ on E, which contradicts $x_{n_0+1}\leqslant x^{\dagger}$ a.e. This negative convergence result for the residual can be explained by the fact that the domain of S is not dense in $L^2(\Omega)$, which consequently does not allow the MHDM to iteratively adjust the approximation with each step. Therefore, any kind of convergence we can hope to achieve will be different from pointwise or (weak) L^2 convergence.

3. Extension of the algorithm with flexible penalty terms

The idea of the MHDM as defined in (1.6) and (1.7) is to look for solutions of (1.3) that show similar behavior on different scales. In this section, we present a more flexible version of the MHDM, which aims to recover solutions with different behavior on different scales. To this end, we introduce a scheme with more general penalty terms. Consider a sequence $(J_n)_{n\in\mathbb{N}_0}$ of functionals on X and define a sequence of approximate solutions by computing

$$u_n\in\operatorname*{arg\,min}_{u\in X}\frac{1}{2}\left\|v_{n-1}-Tu\right\|^2+J_n(u),\tag{3.1}$$

with $x_n=\sum_{k=0}^{n}u_k$, $v_n=f-Tx_n$ and $x_{-1}=0$ as before. In particular, the choice $J_n=\frac{1}{\lambda_n}J$ for some fixed J yields the original MHDM.

In the sequel, we show an interesting norm decomposition of the data, as well as convergence of the residual for this generalized MHDM. Then, we point out a couple of special choices for the functionals J_n which yield known iterative methods for solving (1.2).

3.1. Multiscale norm decomposition of the data

Let us start with a decomposition result for the norm of the data f which, adapted to TV-deblurring, can be found in theorem 2.8 of [2]. Due to the flexibility of the penalty terms considered for this extension, the result can be transferred to related iterative schemes, as we will illustrate below. To simplify notation, we define

$$\zeta_k := T^*(f - Tx_k) = T^* v_k \in \partial J_k(u_k) \tag{3.2}$$

for $k \in \mathbb{N}_0$.

Theorem 3.1. *Let $(J_n)_{n\in\mathbb{N}_0}$ be a sequence of proper, convex, lower-semicontinuous functions such that a sequence $(u_n)_{n\in\mathbb{N}_0}$ is well-defined via (3.1). Then for any $n \in \mathbb{N}_0$, one has*

$$\|f\|^2 = \|v_n\|^2 + \sum_{k=0}^{n} \left(\|Tu_k\|^2 + 2\langle \zeta_k, u_k\rangle \right). \tag{3.3}$$

Proof. For $k \in \mathbb{N}_0$, it is $v_{k-1} = v_k + Tu_k$. Therefore,

$$\|v_{k-1}\|^2 = \|v_k\|^2 + \|Tu_k\|^2 + 2\langle v_k, Tu_k\rangle = \|v_k\|^2 + \|Tu_k\|^2 + 2\langle \zeta_k, u_k\rangle.$$

Telescoping, we obtain

$$\sum_{k=0}^{n} \left(\|Tu_k\|^2 + 2\langle \zeta_k, u_k\rangle \right) = \sum_{k=0}^{n} \left(\|v_{k-1}\|^2 - \|v_k\|^2 \right) = \|f\|^2 - \|v_n\|^2.$$

□

We now extend the class of functionals for which convergence of the residual can be shown by modifying the proof of theorem 2.8 in [2]. In particular, we focus on the case of, possibly different, seminorm penalties. For this, let us recall the characterization of the subgradient of seminorms. For any seminorm J and any $x_0 \in \operatorname{dom} J$, it is (see for instance theorem 2.4.14 in [25]):

$$\partial J(x_0) = \{x^* \in X : \langle x^*, x_0\rangle = J(x_0),\ \langle x^*, x\rangle \leqslant J(x) \text{ for all } x \in X\}. \tag{3.4}$$

Remark. In the case of the original MHDM, where $J_n = (\lambda_n)^{-1} J$ with a seminorm J, we obtain that $\lambda_n \zeta_n \in \partial J(u_n)$, which by (3.4) implies $\langle \zeta_n, u_n\rangle = \frac{J(u_n)}{\lambda_n}$. Thus, (3.3) reads as

$$\|f\|^2 = \|v_n\|^2 + \sum_{k=0}^{n} \left(\|Tu_k\|^2 + 2\frac{J(u_k)}{\lambda_k} \right). \tag{3.5}$$

In particular, (3.3) means

$$\sum_{k=0}^{\infty} \left(\|Tu_k\|^2 + 2\langle \zeta_k, u_k\rangle \right) \leqslant \|f\|^2,$$

with equality if and only if $\|v_n\|$ converges to 0. Theorem 2.1 or [5, Theorem 2.1] give sufficient conditions on the choice of parameters for this convergence to happen.

Lemma 3.2. *Let $(J_n)_{n\in\mathbb{N}_0}$ be a sequence of seminorms and assume there are constants $C_{l,j}\geqslant 0$ such that*

$$J_l(u)\leqslant C_{l,j}J_j(u)$$

for all $u\in X$ and $l,j\in\mathbb{N}_0$ with $l>j$. Furthermore, assume $\lim\limits_{l\to\infty}C_{l,j}=0$ for all $j\in\mathbb{N}_0$. If $J_0(x^\dagger)<\infty$, then the sequence of residuals (v_n) with $v_n=f-Tx_n$ defined by the generalized MHDM method (3.1) converges to 0 in the strong topology of X.

Proof. Fix $k\in\mathbb{N}_0$ and let $N\in\mathbb{N}$. One has

$$\|v_{k+N}\|^2=\langle v_{k+N},v_{k+N}\rangle=\langle v_{k+N},v_k\rangle-\left\langle v_{k+N},\sum_{i=k+1}^{k+N}Tu_i\right\rangle.$$

We will show that both summands on the right-hand side of the former equality converge to 0. For the first one, recall that $T^*v_i\in\partial J_i(u_i)$ by (3.2), which with (3.4) implies $\langle T^*v_i,y\rangle\leqslant J_i(y)$ for all $y\in X$, with equality if and only if $y=u_i$. Additionally, let $\varepsilon>0$. Using the triangle inequality of J_{k+N}, it is

$$\begin{aligned}\langle v_{k+N},v_k\rangle&=\langle v_{k+N},f-Tx_k\rangle=\left\langle T^*v_{k+N},x^\dagger-\sum_{i=0}^k u_i\right\rangle\leqslant J_{k+N}\left(x^\dagger-\sum_{i=0}^k u_i\right)\\&\leqslant J_{k+N}\left(x^\dagger\right)+\sum_{i=0}^k J_{k+N}(u_i)\leqslant C_{k+N,0}J_0\left(x^\dagger\right)+\sum_{i=0}^k C_{k+N,i}J_i(u_i)\\&=C_{k+N,0}J_0\left(x^\dagger\right)+\sum_{i=0}^k C_{k+N,i}\langle T^*v_i,u_i\rangle\leqslant C_{k+N,0}J_0\left(x^\dagger\right)+\frac{\varepsilon}{2}\|f\|^2,\end{aligned}$$

where the last inequality follows from (3.3) if N is chosen large enough such that $C_{k+N,i}\leqslant\varepsilon$ for all $i\leqslant k$. Letting $N\to\infty$ now implies convergence of the first summand. Using the triangle inequality and the subdifferential property of T^*v_{k+N} again, we estimate

$$\left|\left\langle v_{k+N},\sum_{i=k+1}^{k+N}Tu_i\right\rangle\right|=\left|\left\langle T^*v_{k+N},\sum_{i=k+1}^{k+N}u_i\right\rangle\right|\leqslant\sum_{i=k+1}^{k+N}J_{k+N}(u_i)\leqslant\sum_{i=k+1}^{k+N}C_{k+N,k+1}J_i(u_i).$$

Letting $N\to\infty$ and taking the boundedness of $\sum_{k=0}^\infty J_k(u_k)$ by (3.3) into consideration, we obtain that the second summand converges to 0, too. □

Remark. In the situation of the original MHDM, it is $J_l=\frac{1}{\lambda_l}J$, which yields $J_l=\frac{\lambda_j}{\lambda_l}J_j$. Thus, the assumptions of lemma 3.2 are satisfied if and only if $\lim\limits_{j\to\infty}\lambda_j=\infty$ and $J(x^\dagger)<\infty$.

In the case of arbitrary convex penalty terms J_n, we can also adapt part (*ii*) of theorem 2.1.

Lemma 3.3. *Let u_n be obtained by (3.1) with a sequence of proper, convex, lower-semicontinuous functionals $(J_n)_{n\in\mathbb{N}}$. Assume furthermore that there is $C\in\mathbb{R}$ such that $\{u:\limsup\limits_{n\to\infty}J_n(u)<C\}$ is dense in X. If the sequence $(J_n)_{n\in\mathbb{N}_0}$ is uniformly bounded from below and $\sum_{n=0}^\infty J_n(0)<\infty$, then $(f-Tx_n)_{n\in\mathbb{N}_0}$ is bounded, and every weak limit point is in the kernel of T^*. In particular, this implies that $T^*(f-Tx_n)$ converges to 0 in the weak-*-topology of X^*.*

Proof. By the same reasoning as in the proof of theorem 2.1, we obtain

$$\|f - Tx_n\|^2 \leqslant \|f\|^2 + \sum_{k=0}^{n} J_k(0)$$

and

$$\langle T^*(f - Tx_n), z\rangle \leqslant J_n(z) + K$$

for some constant $K \geqslant 0$ and all $z \in X$. Passing to weak limit points of $f - Tx_n$ proves the claim in the same way as in part (*ii*) of theorem 2.1. □

Let us now illustrate how the decomposition result from theorem 3.1 can be applied to related iterative methods.

3.2. The tight MHDM

A tight version of the MHDM was introduced in [5] in order to ensure boundedness, and consequently convergence of the iterates x_n:

$$u_n = \arg\min_{u \in X} \frac{\lambda_n}{2} \|T(u + x_{n-1}) - f\|^2 + \lambda_n a_n J(u + x_{n-1}) + J(u), \quad n \in \mathbb{N}_0, \tag{3.6}$$

with u_n playing the same role as above, i.e. $u_n = x_n - x_{n-1}$. This tight iteration was generalized in [15] to a refined version:

$$u_n = \arg\min_{u \in X} \frac{\lambda_n}{2} \|T(u + x_{n-1}) - f\|^2 + \lambda_n a_n J(u + x_{n-1}) + R_n(u), \quad n \in \mathbb{N}_0, \tag{3.7}$$

where R_n are seminorms. Under mild conditions (see [5, 15]), convergence of the iterates could be proved up to subsequences for both the tight MHDM and its refinement.

Clearly, (3.7) is a special case of the general iteration (3.1) with J_n defined as

$$J_n(u) = \lambda_n a_n J(u + x_{n-1}) + R_n(u).$$

By lemma 3.1, we obtain the following decomposition:

$$\|f\|^2 = \sum_{k=0}^{\infty} \left(\|Tx_k - Tx_{k-1}\|^2 + 2\langle \lambda_k a_k \partial J(x_k) + \partial R_k(u_k), x_k - x_{k-1}\rangle \right),$$

assuming that the conditions for convergence in [15] hold. Here $\partial J(x_k)$ and $\partial R_k(u_k)$ are generic notations for the appropriate subgradients.

3.3. Bregman iteration

The well known Bregman iteration can also be considered in the framework of generalized penalty terms (3.1). Recall that the Bregman iteration is defined via (1.12), i.e.

$$x_n \in \arg\min_{x \in X} \frac{\lambda_n}{2} \|Tx - f\|^2 + D_J^{p_{n-1}}(x, x_{n-1}), \quad n \in \mathbb{N},$$

where for the initial step we set $p_0 = 0$ and afterward choose

$$p_n = \lambda_n T^* (f - Tx_n) + p_{n-1} \in \partial J(x_n).$$

Substituting $u = x - x_{n-1}$ in the definition of x_n and omitting those terms in the definition of the Bregman distance (1.11) which are independent of u, we observe that $u_n := x_n - x_{n-1}$ is a minimizer of (3.1) with

$$J_n(u) := \frac{1}{\lambda_n}\left(J(x_{n-1} + u) - \langle p_{n-1}, u\rangle\right), \qquad p_{n-1} \in \partial J(x_{n-1}).$$

Hence, we may apply the decomposition result (3.3) to the Bregman iteration. Due to

$$\frac{1}{\lambda_n}(p_n - p_{n-1}) \in \partial J(x_n) - \{p_{n-1}\} \in \partial J_n(x_n),$$

we obtain

$$\|f\|^2 = \|Tx_n - f\|^2 + \sum_{k=0}^{n} \left(\|Tu_k\|^2 + 2\lambda_k^{-1} D_J^{sym}(x_k, x_{k-1})\right),$$

with the symmetric Bregman distance

$$D_J^{sym}(x_k, x_{k-1}) = \langle p_k - p_{k-1}, x_k - x_{k-1}\rangle.$$

Since the residual $Tx_n - f$ converges to 0 as $n \to \infty$ for appropriate parameters λ_n (see [13, 14]), we get the full decomposition

$$\|f\|^2 = \sum_{k=0}^{\infty} \left(\|Tx_k - Tx_{k-1}\|^2 + 2\lambda_k^{-1} D_J^{sym}(x_k, x_{k-1})\right).$$

As recalled in the introduction, in the case of $J(u) = \frac{\|u\|^2}{2}$ in Hilbert spaces, the Bregman iteration becomes the iterated Tikhonov regularization (1.9), for which the symmetric Bregman distance $D_J^{sym}(x_k, x_{k-1})$ reads $\|x_k - x_{k-1}\|^2$. Thus, the multiscale decomposition for iterated Tikhonov regularization in Hilbert spaces has the form

$$\|f\|^2 = \sum_{k=0}^{\infty} \left(\|Tx_k - Tx_{k-1}\|^2 + 2\lambda_k^{-1} \|x_k - x_{k-1}\|^2\right).$$

To the best of our knowledge, this result does not seem to be known.

4. Comparison of the MHDM and the generalized Tikhonov regularization

We will now focus on comparing the iterative MHDM to a single step Tikhonov regularization, i.e. classical generalized Tikhonov regularization as defined in (1.5). Generally, the first iterate x_0 of the MHDM is by definition the Tikhonov regularizer at scale λ_0. Nonetheless, we do not expect the MHDM iterate x_k to coincide with the solution x_{λ_k} of the Tikhonov regularization corresponding to the parameter λ_k. Yet, there are frameworks in which this unexpected situation occurs anyway. As an introductory example, consider the case of sparse denoising.

Example 4.1. Let $X = \ell^2$, $T = Id$ and $J = \ell^1$. Using the substitution $x = u + x_{n-1}$ as in (1.7) and setting $x_{-1} = 0$, each iteration step means to compute

$$x_n = \arg\min_{x\in\ell^2}\left\{\frac{\lambda_n}{2}\|x-f\|_{\ell^2}^2 + \|x-x_{n-1}\|_{\ell^1}\right\}.$$

Note that the ith component of the iterate x_n verifies

$$x_n^i = \arg\min_{s\in\mathbb{R}}\left\{\frac{\lambda_n}{2}\left(s-f^i\right)^2 + \left|s-x_{n-1}^i\right|\right\}.$$

Therefore

$$x_n^i = \begin{cases} f^i - \frac{1}{\lambda_n} & \text{if } f^i > x_{n-1}^i + \frac{1}{\lambda_n} \\ f^i + \frac{1}{\lambda_n} & \text{if } f^i < x_{n-1}^i - \frac{1}{\lambda_n} \\ x_{n-1}^i & \text{if } \left|f^i - x_{n-1}^i\right| \leqslant \frac{1}{\lambda_n} \end{cases}. \tag{4.1}$$

Now we can analyze the sequence generated by the MHDM under the assumption that $(\lambda_n)_{n\in\mathbb{N}_0}$ is strictly increasing. We show by induction that for any component i, one has

$$x_n^i = \begin{cases} f^i - \frac{1}{\lambda_n} & \text{if } f^i > +\frac{1}{\lambda_n} \\ f^i + \frac{1}{\lambda_n} & \text{if } f^i < -\frac{1}{\lambda_n} \\ 0 & \text{if } \left|f^i\right| \leqslant \frac{1}{\lambda_n} \end{cases}, \tag{4.2}$$

meaning that performing the first $(n+1)$ iterations of the MHDM is nothing but applying the soft shrinkage operator at scale $\frac{1}{\lambda_n}$. In other words, the $(n+1)$th step of the MHDM procedure is the same as the convex ℓ^2 regularization with parameter λ_n and ℓ^1-penalty term,

$$x_n = \arg\min_{x\in\ell^2}\left\{\frac{\lambda_n}{2}\|x-f\|_{\ell^2}^2 + \|x\|_{\ell^1}\right\}.$$

Indeed, for $n = 0$ this is true by (4.1), since by definition $x_{-1} = 0$. Now, assume that (4.2) holds for some $n \in \mathbb{N}$. We distinguish three cases:

1. Assume $f^i > \frac{1}{\lambda_{n+1}}$. If additionally $f^i > \frac{1}{\lambda_n}$, we must have $x_n^i = f^i - \frac{1}{\lambda_n}$ by assumption and therefore, it is $f^i > f^i - \frac{1}{\lambda_n} + \frac{1}{\lambda_{n+1}} = x_n^i + \frac{1}{\lambda_{n+1}}$, so that (4.1) with n replaced by $n+1$ implies $x_{n+1}^i = f^i - \frac{1}{\lambda_{n+1}}$. Otherwise, it must be $\frac{1}{\lambda_{n+1}} < f^i \leqslant \frac{1}{\lambda_n}$. In that case we have $x_n^i = 0$ and hence $f^i > x_n^i + \frac{1}{\lambda_{n+1}}$. Again, (4.1) with n replaced by $n+1$ yields $x_{n+1}^i = f^i - \frac{1}{\lambda_{n+1}}$.
2. If $f^i < -\frac{1}{\lambda_{n+1}}$, we obtain analogously to the previous case that $x_{n+1}^i = f^i + \frac{1}{\lambda_{n+1}}$.
3. For $\left|f^i\right| \leqslant \frac{1}{\lambda_{n+1}}$, we must also have $\left|f^i\right| \leqslant \frac{1}{\lambda_n}$ by the monotonicity of the λ_n. Therefore, one has $x_n^i = 0$ by assumption, and $\left|f^i - x_n^i\right| = \left|f^i\right| \leqslant \frac{1}{\lambda_{n+1}}$. Thus, (4.1) with n replaced by $n+1$ yields $x_{n+1}^i = 0$.

We will now characterize under which conditions the $(n+1)$th step of the MHDM and the Tikhonov iteration with parameter λ_n coincide. For the remainder of the section, let J be a seminorm. In our further analysis, we consider a dual seminorm which will help to characterize minimizers of the Tikhonov functional (1.4). In the special case where $J = |\cdot|_{\mathrm{TV}}$, the results on the dual norm can be found in [26, 27]. The reader is referred to section 1.3 of [28] for the general case of a seminorm J.

Definition 4.2. The map

$$|\cdot|_* : X^* \to \mathbb{R} \cup \{\infty\}, \, |x^*|_* = \sup_{J(x) \neq 0} \left\langle x^*, \frac{x}{J(x)} \right\rangle$$

is called the dual seminorm of X induced by the seminorm J, where by convention $\frac{x}{J(x)} = 0$ if $J(x) = \infty$.

It can be easily seen that $|\cdot|_*$ is indeed a seminorm.

Remark. The characterization (3.4) implies that $|x^*|_* = 1$, for any x_0 and any subgradient $x^* \in \partial J(x_0)$.

Using the same arguments as in section 2.1 of [2], we can characterize the minimizers of the Tikhonov functional.

Lemma 4.3. *Let J be a seminorm. Let $f \in X$ and $\lambda > 0$. The following statements hold true:*

1. $x_\lambda = 0$ *is a minimizer of* (1.4) *if and only if* $|T^* f|_* \leqslant \frac{1}{\lambda}$.
2. *If* $\frac{1}{\lambda} < |T^* f|_* < \infty$, *then* x_λ *minimizes* (1.4) *if and only if* $|T^*(f - Tx_\lambda)|_* = \frac{1}{\lambda}$ *and* $\langle T^*(f - Tx_\lambda), x_\lambda \rangle = \frac{1}{\lambda} J(x_\lambda)$.

We can now use lemma 4.3 to analyze when the Tikhonov regularization agrees with the MHDM. Let us recall some notation. For $k \in \mathbb{N}_0$, denote by x_{λ_k} a minimizer of the Tikhonov functional with parameter λ_k cf (1.5).

A subgradient of J at x_{λ_k} is given by

$$\xi_{\lambda_k} := \lambda_k T^* \left(f - Tx_{\lambda_k}\right) \in \partial J\left(x_{\lambda_k}\right). \tag{4.3}$$

Analogously, note that the MHDM iterate $x_k = x_{k-1} + u_k$ with u_k as computed in (1.6) can be equivalently obtained as

$$x_k \in \arg\min_{x \in X} \frac{\lambda_k}{2} \|Tx - f\|^2 + J\left(x - x_{k-1}\right), \tag{4.4}$$

with

$$\xi_k := \lambda_k T^* \left(f - Tx_k\right) \in \partial J\left(x_k - x_{k-1}\right). \tag{4.5}$$

In general, a comparison of x_{λ_k} and x_k can be done by using the dual seminorm. For instance, in the case of TV-denoising it was pointed out in section 3.2 of [1] that $\|x_{\lambda_k} - x_k\|_{W^{-1,\infty}} \leqslant \frac{1}{\lambda_k}$. We present next the general form of this result.

Corollary 4.4. *Let J be a seminorm and let x_k be obtained as in* (1.6) *and* (1.7)*. Then, for any $k \in \mathbb{N}_0$, it holds*

$$|T^* T(x_{\lambda_k} - x_k)|_* \leqslant \frac{2}{\lambda_k}. \tag{4.6}$$

Proof. By (4.3), (4.5), (3.4) and the definition of $|\cdot|_*$, one has

$$|T^* \left(f - Tx_k\right)|_* = \left|\frac{\xi_k}{\lambda_k}\right|_* = \frac{1}{\lambda_k} \qquad \text{and} \qquad |T^* \left(f - Tx_{\lambda_k}\right)|_* = \left|\frac{\xi_{\lambda_k}}{\lambda_k}\right|_* = \frac{1}{\lambda_k}.$$

Thus, (4.6) follows by the triangle inequality. □

Let us return to the question of when the MHDM iterates agree with the generalized Tikhonov regularization. Since the first iterate x_0 is obtained via a Tikhonov regularization with parameter λ_0, the base case for an inductive proof holds.

The next theorem verifies the induction step $k \to k+1$. Note that the minimizers of Tikhonov regularization problems (1.4) are not necessarily unique. Therefore, the equality $x_{\lambda_k} = x_k$ should be understood as choosing the same minimizer for both MHDM and Tikhonov minimization problem.

Theorem 4.5. *Let J be a seminorm. Let $(\lambda_k)_{k\in\mathbb{N}}$ be an increasing sequence of positive parameters. Fix $k \in \mathbb{N}$ and assume that $x_{\lambda_k} = x_k$. Then the solution $x_{\lambda_{k+1}}$ of the Tikhonov regularization with parameter λ_{k+1} minimizes the same functional as the MHDM iterate x_{k+1}, i.e.*

$$x_{\lambda_{k+1}} \in \underset{x\in X}{\arg\min} \left\{ \frac{\lambda_{k+1}}{2} \|Tx - f\|^2 + J(x - x_k) \right\}, \tag{4.7}$$

if and only if

$$D_J^{\xi_{\lambda_{k+1}}} \left(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_{k+1}}\right) = 0. \tag{4.8}$$

Proof. First assume that x_{k+1} and $x_{\lambda_{k+1}}$ coincide, i.e. (4.7) holds. This implies $\xi_{\lambda_{k+1}} \in \partial J(x_{\lambda_{k+1}} - x_{\lambda_k})$. We therefore obtain

$$\begin{aligned} 0 \leqslant D_J^{\xi_{k+1}} \left(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_{k+1}}\right) &= J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) - J\left(x_{\lambda_{k+1}}\right) + \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_k}\right\rangle \\ &= -\left(J\left(x_{\lambda_{k+1}}\right) - J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) - \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_k}\right\rangle\right) = -D_J^{\xi_{\lambda_{k+1}}} \left(x_{\lambda_{k+1}}, x_{\lambda_{k+1}} - x_{\lambda_k}\right) \leqslant 0. \end{aligned}$$

Conversely, assume $D_J^{\xi_{\lambda_{k+1}}}(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_{k+1}}) = 0$. We show that $\xi_{\lambda_{k+1}}$ satisfies the optimality conditions of the MHDM, that is $\xi_{\lambda_{k+1}} \in \partial J(x_{\lambda_{k+1}} - x_{\lambda_k})$. Since $\xi_{\lambda_{k+1}} \in \partial J(x_{\lambda_{k+1}})$, it is by (3.4) $\left|\xi_{\lambda_{k+1}}\right|_* = 1$ and we only need to show $\left\langle \xi_{\lambda_{k+1}}, x_{\lambda_{k+1}} - x_{\lambda_k}\right\rangle = J(x_{\lambda_{k+1}} - x_{\lambda_k})$. Indeed, since $x_{\lambda_{k+1}}$ minimizes the Tikhonov functional, we have $\left\langle \xi_{\lambda_{k+1}}, x_{\lambda_{k+1}}\right\rangle = J(x_{\lambda_{k+1}})$. Thus, by (4.8) it is

$$\begin{aligned} \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_{k+1}} - x_{\lambda_k}\right\rangle &= \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_{k+1}} - x_{\lambda_k}\right\rangle + D_J^{\xi_{\lambda_{k+1}}} \left(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_{k+1}}\right) \\ &= \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_{k+1}} - x_{\lambda_k}\right\rangle + J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) - J\left(x_{\lambda_{k+1}}\right) + \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_k}\right\rangle \\ &= \left\langle \xi_{\lambda_{k+1}}, x_{\lambda_{k+1}}\right\rangle + J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) - J\left(x_{\lambda_{k+1}}\right) \\ &= J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right). \end{aligned}$$

□

Remark. If the minimizer of the Tikhonov functional (1.4) is unique, then theorem 4.5 states that $x_{\lambda_k} = x_k$ for all $k \in \mathbb{N}_0$ if and only if $D_J^{\xi_{k+1}}(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_k}) = 0$ for all $k \in \mathbb{N}_0$.

By symmetry, we can also characterize when the iterates of the MHDM minimize the corresponding Tikhonov functional.

Corollary 4.6. *Let $(\lambda_k)_{k\in\mathbb{N}}$ be an increasing sequence of positive parameters. Fix $k \in \mathbb{N}$ and assume that x_k minimizes the Tikhonov functional with parameter λ_k. Then x_{k+1} minimizes the Tikhonov functional with parameter λ_{k+1} if and only if*

$$D_J^{\xi_{k+1}} \left(u_{k+1}, x_{k+1}\right) = 0. \tag{4.9}$$

Proof. The proof follows analogously to the one in theorem 4.5. □

It is remarkable that we can now see whether the iterates of the MHDM can also be obtained via Tikhonov regularization just by knowing the Tikhonov minimizers. We will thus derive an equivalent formulation of (4.8), which is easily verifiable knowing the Tikhonov minimizers. For this, we need to characterize the intersection of subdifferentials of seminorms.

Proposition 4.7. *Let X be a normed space and J be a seminorm on X. For any $z_1, z_2 \in X$ and $z^* \in X^*$, the following equivalence holds:*

$$z^* \in \partial J(z_1) \cap \partial J(z_2) \text{ if and only if } z^* \in \partial J(z_1+z_2) \text{ and } J(z_1)+J(z_2) = J(z_1+z_2).$$

Proof. First assume $z^* \in \partial J(z_1) \cap \partial J(z_2)$. By (3.4) this means $\langle z^*, z\rangle \leqslant J(z)$ for all $z \in X$, $\langle z^*, z_1\rangle = J(z_1)$ and $\langle z^*, z_2\rangle = J(z_2)$. Therefore by the triangle inequality of J and again (3.4), one has

$$J(z_1+z_2) \leqslant J(z_1)+J(z_2) = \langle z^*, z_1+z_2\rangle \leqslant J(z_1+z_2).$$

This implies $\langle z^*, z_1+z_2\rangle = J(z_1+z_2)$, that is $z^* \in \partial J(z_1+z_2)$. Furthermore, we obtain $J(z_1)+J(z_2) = J(z_1+z_2)$.

Conversely, assume $z^* \in \partial J(z_1+z_2)$ and $J(z_1)+J(z_2) = J(z_1+z_2)$. By (3.4), we immediately get $\langle z^*, z\rangle \leqslant J(z)$ for all $z \in X$ and

$$\langle z^*, z_1+z_2\rangle = J(z_1+z_2) = J(z_1)+J(z_2).$$

This yields,

$$J(z_1) = J(z_1+z_2) - J(z_2) = \langle z^*, z_1+z_2\rangle - J(z_2) \leqslant \langle z^*, z_1+z_2\rangle - \langle z^*, z_2\rangle = \langle z^*, z_1\rangle \leqslant J(z_1).$$

Hence, it is $\langle z^*, z_1\rangle = J(z_1)$ and thus $z^* \in \partial J(z_1)$. By analogous reasoning we get $z^* \in \partial J(z_2)$. □

Lemma 4.8. *The condition for agreement of Tikhonov and MHDM, condition* (4.8)*, is equivalent to*

$$J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) + J\left(x_{\lambda_{k+1}}\right) = J\left(2x_{\lambda_{k+1}} - x_{\lambda_k}\right) \tag{4.10}$$

Proof. We have seen in the proof of theorem 4.5 that condition (4.8) is equivalent to

$$\xi_{\lambda_{k+1}} \in \partial J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) \cap \partial J\left(x_{\lambda_{k+1}}\right).$$

Then applying proposition 4.7 proves the claim. □

Let us now verify some examples using (4.10).

Example 4.9. We revisit example 4.1. That is $X = \ell^2$, $J = \|\cdot\|_{\ell^1}$ and $T = Id$. Let now $(\lambda_k)_{k\in\mathbb{N}_0}$ be an increasing sequence of positive numbers. The regularizers x_{λ_k} coincide with the iterates of the MHDM as defined in (4.2). It suffices to verify (4.10) componentwise. Let $k, i \in \mathbb{N}$. We distinguish three cases:

1. Assume $f^i = 0$. Then $x^i_j = 0$ for all $j \in \mathbb{N}$ and (4.10) clearly holds.
2. Assume $f^i > 0$. If $f^i \leqslant \frac{1}{\lambda_{k+1}}$, we have $x^i_{k+1} = 0$ and thus (4.10) holds. Thus assume $f^i > \frac{1}{\lambda_{k+1}}$. Once again we do not need to consider the case $f^i \leqslant \frac{1}{\lambda_k}$. So we assume $f^i \geqslant \frac{1}{\lambda_k}$. This yields $x^i_{\lambda_k} = f^i - \frac{1}{\lambda_k} > 0$ and $x^i_{k+1} = f^i - \frac{1}{\lambda_{k+1}} > 0$. We therefore obtain by the monotonicity of λ_k that

$$\left|x^i_{\lambda_{k+1}} - x^i_{\lambda_k}\right| + \left|x^i_{\lambda_{k+1}}\right| = \frac{1}{\lambda_k} - \frac{1}{\lambda_{k+1}} + f^i - \frac{1}{\lambda_{k+1}} = 2\left(f^i - \frac{1}{\lambda_{k+1}}\right) - \frac{1}{\lambda_k}$$
$$= \left|2x^i_{\lambda_{k+1}} - x^i_{\lambda_k}\right|.$$

3. If $f^i < 0$ the claim follows analogously to the previous case.

Hence, we see that (4.10) holds for all $k \in \mathbb{N}$. Since the initial step of the MHDM is to compute a Tikhonov minimization with parameter λ_0 and $T = Id$ (implying that the Tikhonov minimizers are unique), we can conclude again that the iterates of the MHDM coincide with the Tikhonov minimizers at the corresponding parameter.

The next part deals with an extension of this example to the case of Tikhonov regularization with ℓ^1 penalty for general operators T, where sufficient conditions such that the MHDM iteration coincides with Tikhonov regularization are established.

4.1. ℓ^1-regularization in finite dimensions

The aim of this subsection is to prove that the so-called positive cone condition (see [16]) implies (4.10) in a finite-dimensional case for ℓ^1 regularization. We consider the ℓ^1 Tikhonov regularization

$$\frac{\lambda}{2}\|Tx - y\|_2^2 + \|x\|_1, \tag{4.11}$$

where λ is a positive parameter and $T \in \mathbb{R}^{m\times n}$ is an injective matrix: $\ker T = \{0\}$. Note that this implies that T^*T is invertible. We denote by x_λ the minimizer of (4.11). In the following, the notation x^i is used to refer to the ith component of a vector x.

Remark . One can check that condition (4.10) is verified in case $J = \|.\|_1$ if the inequality

$$\left|x^i_{\lambda_k}\right| \leqslant \left|x^i_{\lambda_{k+1}}\right| \tag{4.12}$$

holds for all components where $x^i_{\lambda_{k+1}} \neq 0$.

We formulate the positive cone condition of [16], stated as diagonal dominance in [29].

Definition 4.10. For some index set $J \subset \{1,\ldots,n\}$, we denote by T^J the corresponding submatrix of T that takes only the columns in J, that is, $(T^J)_{i,j} = T_{i,j}$ for all $i \in \{1,\ldots m\}$ and $j \in J$. Define S^J as the matrix $S_J := \left((T^J)^*T^J\right)^{-1} \in \mathbb{R}^{|J|\times|J|}$.

We say that a matrix T satisfies the *positive cone condition* if for all $J \subset \{1,\ldots,m\}$, the matrix S_J is diagonally dominant, i.e.

$$r_{J,i} := (S_J)_{i,i} - \sum_{j\neq i} |(S_J)_{i,j}| \geqslant 0 \qquad \forall i \in J.$$

To verify the positive cone condition, the following result from [29, lemma 4] is useful.

Lemma 4.11. *In the setting of definition 4.10, a matrix T satisfies the positive cone condition if and only if $(T^*T)^{-1}$ is diagonally dominant, i.e. $S := (T^*T)^{-1}$ satisfies*

$$S_{i,i} - \sum_{j\neq i} |S_{i,j}| \geqslant 0 \qquad \forall i \in \{1,\ldots,m\}.$$

Before stating the main result of this subsection, we need some lemmas, and here we strongly rely on the finite-dimensionality.

For later references we state now the optimality conditions for (4.11),

$$\lambda T^*Tx_\lambda + \xi_\lambda = T^*y, \qquad \text{with } \xi_\lambda \in \partial(\|\cdot\|_1)(x_\lambda), \tag{4.13}$$

and $\xi_\lambda^i \in [-1,1]$ with $\xi_\lambda^i = \text{sign}(x_\lambda^i)$ whenever $x_\lambda^i \neq 0$.

We start with a lemma that establishes continuity of the solution x_λ with respect to λ in the finite-dimensional setup (see similar results in [16] under the strict positive cone condition).

Lemma 4.12. *Let $\lambda > 0$. The mapping*

$$\lambda \mapsto x_\lambda$$

is continuous from $\mathbb{R}^+$ into $\mathbb{R}^n$.

Proof. Let $\lambda, \mu \in \mathbb{R}^+$ and subtract (4.13) for λ and μ from each other. This yields

$$T^*T(x_\lambda - x_\mu) + \lambda^{-1}(\xi_\lambda - \xi_\mu) + \left(\lambda^{-1} - \mu^{-1}\right)\xi_\mu = 0.$$

Taking the inner product with $(x_\lambda - x_\mu)$ gives

$$\|T(x_\lambda - x_\mu)\|^2 + \lambda^{-1}\langle \xi_\lambda - \xi_\mu, x_\lambda - x_\mu\rangle = -\left(\lambda^{-1} - \mu^{-1}\right)\langle \xi_\mu, x_\lambda - x_\mu\rangle.$$

By convexity, the second term on the left-hand side is nonnegative, while by injectivity and finite-dimensionality the first term has a lower bound $c\|x_\lambda - x_\mu\|^2$, where c is the smallest singular value of T. Thus,

$$c\|x_\lambda - x_\mu\|^2 \leqslant -\left(\lambda^{-1} - \mu^{-1}\right)\langle \xi_\mu, x_\lambda - x_\mu\rangle \leqslant |\lambda^{-1} - \mu^{-1}|\|x_\lambda - x_\mu\|,$$

which yields continuity. □

For the next results we need the following index sets:

$$N(\lambda) = \left\{i \in \{1,\ldots,n\} : x_\lambda^i = 0\right\}$$
$$I(\lambda) = \left\{i \in \{1,\ldots,n\} : x_\lambda^i \neq 0\right\}.$$

Lemma 4.13. *Let $\mu < \lambda$ and suppose that T satisfies the positive cone condition. Furthermore, suppose that the following conditions hold:*

$$N(\lambda) \subset N(\mu), \tag{4.14}$$

$$\xi_\lambda^i = \xi_\mu^i \qquad \textit{for all } i \in I(\lambda) \tag{4.15}$$

Then

$$\left(x_\mu^i\right) \operatorname{sign}\left(x_\lambda^i\right) \leqslant \left|x_\lambda^i\right| \qquad \textit{for all } i \in I(\lambda). \tag{4.16}$$

Proof. We use a partitioning of the vectors into the index sets $I(\lambda)$ and $N(\lambda)$,

$$x = \begin{bmatrix} x^{I(\lambda)} \\ x^{N(\lambda)} \end{bmatrix},$$

where the upper part contains the indices of $I(\lambda)$ and the lower part those in $N(\lambda)$. Let λ, μ satisfy the conditions above. By using that $x_\mu^i = 0$ for $i \in N(\lambda)$, it follows from the optimality conditions that

$$T^*T \begin{bmatrix} x_\mu^{I(\lambda)} - x_\lambda^{I(\lambda)} \\ 0 \end{bmatrix} + \begin{bmatrix} \left(\mu^{-1} - \lambda^{-1}\right) \xi_\lambda^{I(\lambda)} \\ \mu^{-1}\xi_\mu - \lambda^{-1}\xi_\lambda^{N(\lambda)} \end{bmatrix} = 0,$$

where $\xi_\lambda^{I(\lambda)} := (\xi_\lambda^i)_{i \in I(\lambda)}$ and $\xi_\lambda^{N(\lambda)} := (\xi_\lambda^i)_{i \in N(\lambda)}$. Let $i \in I(\lambda)$. Since $\xi_\lambda^i = \operatorname{sign}(x_\lambda^i)$, we obtain

$$\begin{aligned} x_\mu^i - x_\lambda^i &= -\left(\mu^{-1} - \lambda^{-1}\right)\left(S_{I(\lambda)}\xi_\lambda^{I(\lambda)}\right)_i \\ &= -\left(\mu^{-1} - \lambda^{-1}\right) \operatorname{sign}\left(x_\lambda^i\right) \left[\left(S_{I(\lambda)}\right)_{i,i} + \sum_{j \in I(\lambda)\setminus\{i\}} \left(S_{I(\lambda)}\right)_{i,j} \frac{\operatorname{sign}\left(x_\lambda^j\right)}{\operatorname{sign}\left(x_\lambda^i\right)}\right] \end{aligned}$$

Because

$$\left(S_{I(\lambda)}\right)_{i,i} + \sum_{j \in I(\lambda)\setminus\{i\}} \left(S_{I(\lambda)}\right)_{i,j} \frac{\operatorname{sign}\left(x_\lambda^j\right)}{\operatorname{sign}\left(x_\lambda^i\right)} \geqslant \left(S_{I(\lambda)}\right)_{i,i} - \sum_{j \in I(\lambda)\setminus\{i\}} \left|\left(S_{I(\lambda)}\right)_{i,j}\right| = r_{I(\lambda),i} \geqslant 0,$$

it follows that

$$\left(x_\mu^i - x_\lambda^i\right) \operatorname{sign}\left(x_\lambda^i\right) \leqslant 0,$$

which is equivalent to (4.16) □

Lemma 4.14. *Let $\mu < \lambda$ and suppose that T satisfies the positive cone condition. Then there exists an $\varepsilon > 0$ such that*

$$N(\lambda) \subset N(\mu) \qquad \forall \mu \in [\lambda - \varepsilon, \lambda].$$

Proof. Suppose this is not the case. Then there exist a monotonically increasing sequence $(\mu_k)_{k\in\mathbb{N}}$ with $\mu_k < \lambda$ and $\lim_{k\to\infty} \mu_k = \lambda$, and an index sequence $(i_k)_{k\in\mathbb{N}} \subset N(\lambda)$ such that $x^{i_k}_{\mu_k} \neq 0$ while $x^{i_k}_{\lambda} = 0$. By the continuity in lemma 4.12, this sequence satisfies $\lim_{k\to\infty} x^{i_k}_{\mu_k} = 0$.

We proceed by constructing a subsequence of x_{μ_k} that has the same set of non-zero components for all its elements. Note that by finite dimensionality and continuity (lemma 4.12) we may assume $I(\mu_k) \subset I(\lambda)$ for all k. Now, pick $j \in N(\lambda)$ such that $x^j_{\mu_k} \neq 0$ for infinitely many k. Such an index j must exist since $(i_k)_{k\in\mathbb{N}}$ is a sequence in a finite set, and thus it has to meet some index j infinitely many times. Take a subsequence (again denoted by the index k) such that $x^j_{\mu_k} \neq 0$ for all $k \in \mathbb{N}$, then set $N^\infty = \{j\}$ and $N^F = N(\lambda) \setminus \{j\}$. We now proceed inductively with this construction for $j \in N^F$. Try to find an index $j \in N^F$ such that $x^j_{\mu_k} \neq 0$ for infinitely many k. Such an index may or may not exist. If it exists, we again take a subsequence (again denoted by an index k) such that $x^j_{\mu_k} \neq 0$ for all $k \in \mathbb{N}$, we add j to N^∞ and remove it from N^F. We proceed with this construction until either no such j can be found (case (i)) or N^F is the empty set (case (ii)) - this situation must happen since the index set is a finite set. In case (i), we have for all $j \in N_F$, that $x^j_{\mu_k} \neq 0$ for only finitely many k. Thus, we can take a subsequence (again denoted by an index k) such that $x^j_{\mu_k} = 0$ for all k and all $j \in N^F$. By this construction, we find a subsequence and an index set partition $N^F \cup N^\infty = N(\lambda)$ (including the case of N^F being empty, case (ii)) such that

$$x^j_{\mu_k} \neq 0 \quad \text{for } j \in N^\infty, \qquad x^j_{\mu_k} = 0 \text{ for } j \in N^F.$$

Thus, we have $I(\mu_k) = I(\lambda) \cup N^\infty$ for all k.

We note that the optimality condition and lemma 4.12 also imply that ξ_μ is continuous in μ. For $j \in N^\infty$, we have that $\xi^j_{\mu_k} = \text{sign}(x_{\mu_k}) \in \{-1, 1\}$ is continuous in μ and hence, for k sufficiently large, this implies that the sign remains constant for sufficiently large $k \geqslant n_0$, that is, $\xi^j_{\mu_k} = \xi^j_{\mu_{k'}}$ for all $k' \geqslant k$. Again by continuity it follows for $i \in I(\lambda)$ that $\lim_{k\to\infty} \xi^i_{\mu_k} = \xi^i_\lambda \in \{-1, 1\}$, and hence $\text{sign}(x^i_{\mu_k}) = \xi^i_\mu$ must remain constant also on $I(\lambda)$ for k sufficiently large.

It follows that the constructed sequence satisfies the properties (4.14) and (4.15) for any pair $\mu_k < \mu_{k'}$ with $n_0 \leqslant k \leqslant k'$. Again by continuity we also have $\text{sign}(x^i_{\mu_k}) = \text{sign}(x^i_{\mu_{k'}})$ for $i \in N^\infty$ which implies, using (4.16),

$$|x^i_{\mu_k}| = x^i_{\mu_k} \xi^i_{\mu_k} = x^i_{\mu_{k'}} \xi^i_{\mu_k} \leqslant |x^i_{\mu_{k'}}|, \qquad n_0 \leqslant k \leqslant k', i \in N^\infty.$$

However this contradicts the condition that $x^i_{\mu_{k'}} \to 0$ for $i \in N^\infty$, hence the proposition is verified. □

We prove in the sequel a local monotonicity result.

Proposition 4.15. *Let T satisfy the positive cone condition. Then, for any λ, there exists an $\varepsilon > 0$ such that*

$$|x^i_\mu| \leqslant |x^i_\lambda| \qquad \forall i \in I(\lambda), \forall \mu \in [\lambda - \varepsilon, \lambda]$$

Proof. Let $\mu < \lambda$. It follows by continuity that $\text{sign}(x^i_\lambda) = \xi^i_\lambda = \xi^i_\mu = \text{sign}(x^i_\mu)$ for $i \in I(\lambda)$ and $|\mu - \lambda|$ small enough. Taking additionally μ close enough to λ so that lemma 4.14 applies, we may use lemma 4.13 to conclude the result

$$|x^i_\mu| = x^i_\mu \,\text{sign}\left(x^i_\mu\right) = x^i_\mu \,\text{sign}\left(x^i_\lambda\right) \leqslant |x^i_\lambda|.$$

□

By continuity, we may globalize the result as follows.

Proposition 4.16. *Let T satisfy the positive cone condition. Then, for any $\lambda > 0$, we have*

$$|x^i_\mu| \leqslant |x^i_\lambda| \qquad \forall i \in I(\lambda), \forall \mu \leqslant \lambda.$$

Proof. Suppose that the statement does not hold. Then for some $i \in I(\lambda)$ there exists $\mu^* < \lambda$ with

$$|x^i_{\mu^*}| > |x^i_\lambda|.$$

Take $\bar{\mu}$ as the supremum of all such $\mu^* < \lambda$. It follows then by continuity that $|x^i_{\bar{\mu}}| = |x^i_\lambda|$, and there exists a sequence $(\mu_k)_k$ converging to $\bar{\mu}$ with $\mu_k < \bar{\mu}$ and $|x^i_{\mu_k}| > |x^i_{\bar{\mu}}|$, which contradicts proposition 4.15. □

We note that monotonicity of components was also proven by Meinshausen [30] under the slightly stronger restricted positive cone condition. The result of the theorem has also been stated in [29, remark 3] though without full proof.

We are now in a position to state the main result of this subsection, its proof being a consequence of the auxiliary results shown above.

Theorem 4.17. *Let $T \in \mathbb{R}^{m\times n}$ be injective and assume that T satisfies the positive cone condition. Then the ℓ^1-minimizers of (4.11) satisfy (4.10) with $J = \|\cdot\|_1$ and with corresponding regularization parameters λ_k, for any $k \in \mathbb{N}_0$. In particular, in this case the MHDM iteration agrees with the corresponding ℓ^1-regularization, that is*

$$x_n = x_{\lambda_n}, \forall n \in \mathbb{N}_0.$$

Proof. Proposition 4.15 applies for $\mu = \lambda_k$ and $\lambda = \lambda_{k+1}$ yielding (4.12) and thus, (4.10). □

We finish with a denoising example where the MHDM does not agree with Tikhonov regularization. This will be due to a violation of (4.10).

Example 4.18. Let

$$T = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

and $f = (4, -1)^T$. The corresponding Tikhonov minimizer

$$\left(x^1_\lambda, x^2_\lambda\right) \in \underset{(x_1,x_2)\in\mathbb{R}^2}{\arg\min} \frac{\lambda}{2}\left\|T(x_1,x_2)^T - f\right\|_2^2 + |(x_1,x_2)|_1$$

is given by

$$\left(x_\lambda^1, x_\lambda^2\right) = \begin{cases} (0,0) & \text{for } 0 < \lambda \leqslant \frac{1}{7}, \\ \left(\frac{7}{5} - \frac{1}{5\lambda},\, 0\right) & \text{for } \frac{1}{7} < \lambda \leqslant \frac{1}{2}, \\ \left(\frac{1}{\lambda} - 1,\, 6 - \frac{3}{\lambda}\right) & \text{for } \frac{1}{2} \leqslant \lambda \leqslant 1, \\ \left(0,\, 4 - \frac{1}{\lambda}\right) & \text{for } 1 \leqslant \lambda \leqslant 3, \\ \left(\frac{3}{\lambda} - 1,\, 6 - \frac{7}{\lambda}\right) & \text{for } \lambda \geqslant 3. \end{cases}$$

Note that for $\lambda \in \left(\frac{1}{2}, 1\right)$ we have that x_λ^1 is positive and strictly decreasing to 0, while x_λ^2 is positive and strictly increasing with respect to λ. If the sequence $(\lambda_k)_{k\in\mathbb{N}_0}$ is chosen such that $\lambda_k, \lambda_{k+1} \in \left(\frac{1}{2}, 1\right)$ and $x^1_{\lambda_{k+1}} < x^1_{\lambda_k} < 2x^1_{\lambda_{k+1}}$ hold for some $k \geqslant 0$, we obtain

$$\left\|x_{\lambda_{k+1}} - x_{\lambda_k}\right\|_1 + \left\|x_{\lambda_{k+1}}\right\|_1 = \left|x^1_{\lambda_k}\right| + 2\left|x^2_{\lambda_{k+1}}\right| - \left|x^2_{\lambda_k}\right|$$

but

$$\left\|2x_{\lambda_{k+1}} - x_{\lambda_k}\right\|_1 = 2\left|x^1_{\lambda_{k+1}}\right| - \left|x^1_{\lambda_k}\right| + 2\left|x^2_{\lambda_{k+1}}\right| - \left|x^2_{\lambda_k}\right|.$$

Therefore, condition (4.10) is not satisfied and the MHDM does not agree with Tikhonov minimization. Furthermore, note that

$$(T^*T)^{-1} = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$$

is not diagonally dominant, meaning that T does not satisfy the positive cone condition.

4.2. TV-denoising in one dimension

The results of the previous section can be used to analyze also the one-dimensional TV denoising problem. The main idea is to first consider the finite dimensional problem. By the use of a substitution, we will transform TV-denoising into a problem of the form (4.11), which satisfies the assumptions of theorem 4.17. Note that a similar approach was used in [29, section IV]. The main point is that finite dimensional TV denoising is equivalent to TV denoising on the class of piecewise constant functions with jumps at predetermined points. This will allow using approximation arguments to obtain the infinite dimensional case.

In finite dimensional TV regularization, the penalty is essentially the ℓ^1-norm of the derivative. We consider now the Tikhonov functional

$$\frac{\lambda}{2}\|x - y\|_2^2 + \|Dx\|_1 \tag{4.17}$$

where $y, x \in \mathbb{R}^n$, and we denote by x_λ a minimizer of this functional with respect to x. We consider a finite-dimensional situation and Dx a standard difference quotient, that is, for $x \in \mathbb{R}^n$, the matrix $D \in \mathbb{R}^{(n-1)\times n}$ has the form

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}. \tag{4.18}$$

Note that D has as nullspace $N(D)$, the subspace of constant vectors. Since we can decompose the $\mathbb{R}^n$-space into $N(D) \oplus N(D)^\perp$ orthogonally, it is not difficult to see that (4.17) can be replaced by the corresponding optimization problem with $x \in N(D)^\perp$ and y replaced by $P_{N(D)^\perp} y$, i.e. the orthogonal projection onto $N(D)^\perp$. (The component of x in $N(D)$ is easily calculated as the orthogonal projection of y to $N(D)$.) Thus, by now considering (4.17) in $N(D)^\perp = R(D^*)$, we may set $x = D^* w$ and minimize over $w \in \mathbb{R}^{n-1}$. Upon setting $z = DD^* w$ and observing that DD^* is invertible, we arrive at the function

$$\frac{\lambda}{2} \|D^* (DD^*)^{-1} z - y\|^2 + \|z\|_1$$

to be minimized over z. This is the setup of the previous section with $T = D^*(DD^*)^{-1}$ and

$$(T^*T)^{-1} = DD^* = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix},$$

where the latter operator corresponds to the standard second-order difference quotient. Thus, it satisfies the positive cone condition according to lemma 4.11. As a consequence, the MHDM agrees with Tikhonov regularization in this case by theorem 4.17. Note that this is not necessarily true for a general 1D regularization of the form

$$\frac{\lambda}{2} \|Tx - y\|_2^2 + \|Dx\|_1.$$

with a more general operator T.

By a limit argument, we can prove the same result in the continuous case in one dimension.

Theorem 4.19. *Let $y \in L^2([0,1])$ and consider the TV-denoising problem in one-dimension on the interval $(0,1)$, i.e. (1.4) with $T = Id$ and $J = |.|_{TV}$. Then the MHDM iteration agrees with the corresponding Tikhonov regularization as in theorem 4.5*

Proof. Due to standard density arguments in $L^2([0,1])$, y can be approximated by piecewise constant functions y_N such that $\|y - y_N\|_{L^2(0,1)} \to 0$ as $N \to \infty$. To specify the notation, the function y_N is constant on the intervals (s_i, s_{i+1}), where $(s_i)_{i=0}^N$ represent the nodes on a uniform grid of $[0,1]$.

According to [31, lemma 4.34], the corresponding minimizer of TV-denoising, now denoted by $x_{\lambda_k}^N$, is again piecewise constant on the same grid. One can see that the vectors of coefficients of $x_{\lambda_k}^N$ on (s_i, s_{i+1}) are the solutions of the discrete TV-denoising problems (4.17) with corresponding vector of coefficients of y_N as data y. Moreover, the expression $\|Dx\|_1$ equals the TV-seminorm $|x|_{TV}$ for all piecewise constant x. Note that the MHDM algorithm iteratively provides solutions of denoising problems, and the solutions to those are piecewise constant on the same grid. Consequently, we inductively obtain that applying the MHDM with data y_N is equivalent to applying the MHDM to a discrete denoising problem with the coefficients of y_N as data. Hence, the MHDM agrees with Tikhonov regularization (i.e. TV-denoising) in this case when y is replaced by y_N. We now verify (4.8) by taking limits and using stability of the regularization scheme. For fixed parameter λ_k, $x_{\lambda_k}^N$ depends continuously of y_N in a sense made precise in [9]; see also [32, theorem 2.4]. In fact, the following hold for $N \to \infty$ and for a subsequence (denoted the same as the original sequence):

$$|x_{\lambda_k}^N|_{TV} \to |x_{\lambda_k}|_{TV} \qquad \text{by [9, theorem 3.2],} \tag{4.19}$$

$$x_{\lambda_k}^N \rightharpoonup_{BV}^* x_{\lambda_k} \qquad \text{by [9, theorem 3.2],}$$

$$x_{\lambda_k}^N - y_N \rightharpoonup_{L^2} x_{\lambda_k} - y \qquad \text{see [9, proof of theorem 3.2],} \tag{4.20}$$

$$\|x_{\lambda_k}^N - y_N\| \to \|x_{\lambda_k} - y\| \qquad \text{see [9, equation (8)].} \tag{4.21}$$

Since from any subsequence one can extract another subsequence with those convergence properties, they must hold for the sequence itself. The last two identities imply strong L^2-convergence of $x_{\lambda_k}^N - y_N$ to $x_{\lambda_k} - y$ by the Radon–Riesz property of L^2 (see also [32, equation (12)]). Based on the convergence of y_N to y in $L^2([0,1])$, we can also conclude the convergence of $x_{\lambda_k}^N$ to x_{λ_k}. Since by definition it is $\xi_{\lambda_k} = y - x_{\lambda_k}$, this means that $\xi_{\lambda_k}^N \to \xi_{\lambda_k}$ strongly in $L^2([0,1])$, which yields

$$\left\langle \xi_{\lambda_k}^N, x_{\lambda_k}^N \right\rangle \to \left\langle \xi_{\lambda_k}, x_{\lambda_k} \right\rangle \tag{4.22}$$

for $N \to \infty$. Now consider (4.8) with $J = |.|_{TV}$,

$$D_J\left(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_{k+1}}\right) = J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) - J\left(x_{\lambda_k}\right) - \left\langle \xi_{\lambda_k}, -x_{\lambda_k} \right\rangle.$$

Taking into account that $D_J(x_{\lambda_{k+1}}^N - x_{\lambda_k}^N, x_{\lambda_{k+1}^N}) = 0$ by the results for discrete TV-denoising from above and using the weak lower semicontinuity of J, one obtains

$$\begin{aligned} J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) \leqslant \liminf_{N\to\infty} J\left(x_{\lambda_{k+1}}^N - x_{\lambda_k}^N\right) &= \liminf_{N\to\infty}\left(J\left(x_{\lambda_k}^N\right) + \left\langle \xi_{\lambda_k}^N, -x_{\lambda_k}^N \right\rangle\right) \\ &= J\left(x_{\lambda_k}\right) - \left\langle \xi_{\lambda_k}, x_{\lambda_k} \right\rangle, \end{aligned}$$

where the last equality holds by (4.19) and (4.22). Thus, the following (non-negative) Bregman distance satisfies

$$D_J\left(x_{\lambda_{k+1}} - x_{\lambda_k}, x_{\lambda_{k+1}}\right) = J\left(x_{\lambda_{k+1}} - x_{\lambda_k}\right) - J\left(x_{\lambda_k}\right) + \left\langle \xi_{\lambda_k}, x_{\lambda_k} \right\rangle \leqslant 0,$$

meaning that it must be 0 and implying that (4.8) is verified. □

4.3. TV-denoising in higher dimensions

The previous results that the MHDM iteration agrees with Tikhonov regularization for denoising in the one-dimensional case cannot be extended to the higher-dimensional situation, not even in a discrete case. For TV-denoising on domains in $\mathbb{R}^2$ and $\mathbb{R}^3$ in a finite dimensional framework, the matrix D represents a discretization of the gradient operator $D \sim \nabla$, while DD^* is a discrete version of the operator ∇ div, which does not necessarily satisfy the diagonal dominance condition of lemma 4.11. Thus, in general, the MHDM iteration does not agree with the Tikhonov regularization in higher dimensions. We can actually provide a counterexample to condition (4.10) in the two-dimensional case.

Example 4.20. Let us consider the denoising problem in two dimensions (1.5) with T being the identity. Let X be the space of $L^2(\mathbb{R}^2)$ functions u with bounded (isotropic) TV-seminorm $J(u) < \infty$, where $J = |.|_{TV}$ on $\mathbb{R}^2$. We consider data y given by the characteristic function of the unit square: $y = \chi_{[-1,1]^2}$. In this case, the explicit form of minimizers to (1.5) are known, and they have as level sets 'rounded' squares, i.e. a square whose edges are rounded by circular arcs; see, e.g. [33–35]. Based on this, a useful explicit functional form has been stated in

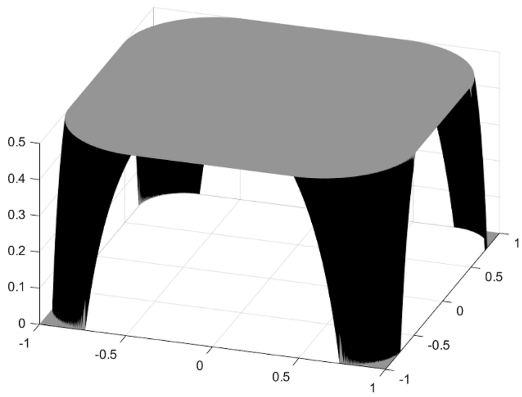

Figure 1. Illustration of x_λ in example 4.20. The function at $(s,t) \in E_\lambda$ is marked in black.

[36, p 1273], as described below. Namely, the minimizers of (1.5) for $\lambda^{-1} < 1/(1+\sqrt{\pi}/2)$ are defined as follows:

$$x_\lambda(s,t) := \begin{cases} 1 - \lambda^{-1}(1+\sqrt{\pi}/2) & r(s,t) \geqslant 1/(1+\sqrt{\pi}/2) \\ 1 - \frac{\lambda^{-1}}{r(s,t)} & \lambda^{-1} < r(s,t) < 1/(1+\sqrt{\pi}/2) \\ 0 & r \leqslant \lambda \text{ or } (s,t) \notin [0,1]^2, \end{cases} \tag{4.23}$$

where $r(s,t) := 2 - |s| - |t| + \sqrt{2(1-|s|)(1-|t|)}$. Of particular interest for us is the region

$$E_\lambda := \left\{(s,t) \in (0,1)^2 \mid \lambda^{-1} < r(s,t) < 1/\left(1+\sqrt{\pi}/2\right)\right\},$$

which is bounded by two circular arcs and parts of the boundary of the unit square, where the solution is smooth. An illustration of x_λ is given in figure 1, where the graph of x_λ at E_λ is marked in black.

Now consider two solutions $x_{\lambda_{k+1}}$ and x_{λ_k} with $\lambda_{k+1} > \lambda_k > (1+\sqrt{\pi}/2)$, and take a fixed small ball B_ε with closure inside the region E_{λ_k}, which is then also included in $E_{\lambda_{k+1}}$. It follows that

$$\nabla x_{\lambda_{k+1}} = -\lambda_{k+1}^{-1} \nabla \frac{1}{r(s,t)} \quad \text{and} \quad \nabla x_{\lambda_k} = -\lambda_k^{-1} \nabla \frac{1}{r(s,t)}, \qquad \text{for } (s,t) \in B_\varepsilon.$$

By smoothness, the TV-norm equals the L^1-norm of the gradient in B_ε, and it holds

$$\begin{aligned} &\int_{B_\varepsilon} |\nabla x_{\lambda_{k+1}} - \nabla x_{\lambda_k}| \mathrm{d}(s,t) + \int_{B_\varepsilon} |\nabla x_{\lambda_{k+1}}| \mathrm{d}(s,t) \\ &\quad = |\lambda_{k+1}^{-1} - \lambda_k^{-1}| \int_{B_\varepsilon} \left|\nabla \frac{1}{r(s,t)}\right| \mathrm{d}(s,t) + |\lambda_{k+1}^{-1}| \int_{B_\varepsilon} \left|\nabla \frac{1}{r(s,t)}\right| \mathrm{d}(s,t) \\ &\quad = \left(|\lambda_{k+1}^{-1} - \lambda_k^{-1}| + |\lambda_{k+1}^{-1}|\right) \int_{B_\varepsilon} \left|\nabla \frac{1}{r(s,t)}\right| \mathrm{d}(s,t) \\ &\int_{B_\varepsilon} |2\nabla x_{\lambda_{k+1}} - \nabla x_{\lambda_k}| \mathrm{d}(s,t) = \left(|2\lambda_{k+1}^{-1} - \lambda_k^{-1}|\right) \int_{B_\varepsilon} \left|\nabla \frac{1}{r(s,t)}\right| \mathrm{d}(s,t). \end{aligned}$$

Since $\lambda_{k+1}^{-1} < \lambda_k^{-1}$, it follows that

$$\left(|\lambda_{k+1}^{-1} - \lambda_k^{-1}| + |\lambda_{k+1}^{-1}|\right) = \lambda_k^{-1} > |2\lambda_{k+1}^{-1} - \lambda_k^{-1}|,$$

and thus the left-hand side in the above identity is strictly larger than the right one:

$$\int_{B_\varepsilon} |\nabla x_{\lambda_{k+1}} - \nabla x_{\lambda_k}| \mathrm{d}(s,t) + \int_{B_\varepsilon} |\nabla x_{\lambda_{k+1}}| \mathrm{d}(s,t) > \int_{B_\varepsilon} |2\nabla x_{\lambda_{k+1}} - \nabla x_{\lambda_k}| \mathrm{d}(s,t).$$

As the solutions $x_{\lambda_{k+1}}$ and x_{λ_k} are smooth in a neighborhood of B_ε, we may decompose the TV-norm (see [37, corollary 3.89]) as

$$|x_{\lambda_{k+1}}|_{TV} = \int_{B_\varepsilon} |\nabla x_{\lambda_{k+1}}| \mathrm{d}(s,t) + |x_{\lambda_{k+1}}|_{\mathrm{TV}(\mathbb{R}^2 \setminus B_\varepsilon)}.$$

One can proceed analogously for x_{λ_k} and the combinations $x_{\lambda_{k+1}} - x_{\lambda_k}$ and $2x_{\lambda_{k+1}} - x_{\lambda_k}$. Now considering (4.10), it follows from the triangle inequality that

$$|x_{\lambda_{k+1}} - x_{\lambda_k}|_{\mathrm{TV}(\mathbb{R}^2 \setminus B_\varepsilon)} + |x_{\lambda_{k+1}}|_{\mathrm{TV}(\mathbb{R}^2 \setminus B_\varepsilon)} \geqslant |2\nabla x_{\lambda_{k+1}} - \nabla x_{\lambda_k}|_{TV(\mathbb{R}^2 \setminus B_\varepsilon)},$$

such that we arrive at

$$|x_{\lambda_{k+1}} - x_{\lambda_k}|_{\mathrm{TV}} + |x_{\lambda_{k+1}}|_{\mathrm{TV}} > |2\nabla x_{\lambda_{k+1}} - \nabla x_{\lambda_k}|_{TV},$$

implying that (4.10) does not hold. Consequently, the MHDM iteration is not identical to Tikhonov regularization in this situation.

Note however, that the set E_λ that yields a violation of (4.10) is rather narrow, such that the difference between the approximate solutions provided by the two methods might be small. In fact, numerical experiments for this setup have only indicated a difference of less than 2% (in the L^2-norm).

On the other hand, for very special data y, it is the case that two-dimensional TV-denoising agrees with the MHDM iteration, namely when y is the characteristic function of so-called calibrable sets, as explained below.

Example 4.21. Let $\Omega \subset \mathbb{R}^2$ be bounded with Lipschitz boundary. We consider the case of TV denoising, i.e. $X = L^2(\Omega)$, $J = |\cdot|_{\mathrm{TV}}$ and $T = Id$. Let $C \subset \Omega$ be a convex set. We furthermore assume that C has $C^{1,1}-$boundary and that the curvature $\kappa_{\partial C}$ of C satisfies $\|\kappa_{\partial C}\|_{L^\infty} \leqslant q(C)$, where $q(C) = \frac{|C|_{\mathrm{TV}}}{m(C)}$, with $m(C)$ being the Lebesgue measure of C. By lemma 4 in [38], this is equivalent to C being convex and having the property that there is a an $\varepsilon > 0$, such that C is the (possibly uncountable) union of balls with radius ε. Define $f = b\,q(C)\chi_C$ for some $b \in \mathbb{R}$. By theorem 4 and proposition 7 in [38], the minimizer of the denoising problem is given by $x_\lambda = \mathrm{sign}(b) \max\left\{b - \frac{1}{\lambda}), 0\right\} q(C)\chi_C$, that is

$$x_\lambda = \begin{cases} \left(b - \frac{1}{\lambda}\right) q(C)\chi_C & \text{if } b > +\frac{1}{\lambda} \\ \left(b + \frac{1}{\lambda}\right) q(C)\chi_C & \text{if } b < -\frac{1}{\lambda} \\ 0 & \text{if } |b| \leqslant \frac{1}{\lambda}. \end{cases}$$

Therefore, relation (4.10) holds by the same reasoning as in example 4.9.

Table 1. Comparison of MHDM iterates and Tikhonov minimizers for one-dimensional TV-regularization.

k	Denoising error e_n Tik.-MHDM	Deblurring error e_n Tik.-MHDM
1	0.0010	0.0021
2	0.0005	0.0799
3	0.0005	0.1155
5	0.0001	0.1179
7	2.8032×10^{-6}	0.1118
11	2.8284×10^{-10}	0.1093

5. Examples and numerical results[3]

In this section, we provide examples of possible penalty functionals to be used for the MHDM. We will focus on the comparison of the MHDM iteration to the generalized Tikhonov regularization.

5.1. *One dimensional TV-regularization*

Let us first investigate the case of one-dimensional TV-regularization, which means considering the functional (1.4) on $L^2([0,1])$ with $J = |\cdot|_{\mathrm{TV}}$. We employ a discretization of the interval $[0,1]$ into $N = 100$ equidistant nodes and approximate the TV via the discrete derivative operator, that is, $|u|_{\mathrm{TV}} \approx \frac{1}{N}\|Du\|_1$ with D as in (4.18). The ground truth is given as a piecewise constant signal $x^\dagger = \chi_{[0.3,0.5]} + \frac{1}{2}\chi_{[0.68,0.72]}$.

We deal with the cases of denoising, i.e. $T = Id$, and deblurring, where T is a convolution operator of a centered Gaussian kernel with standard deviation $\sigma = 0.1$. The solutions for Tikhonov regularization and MHDM were computed using the primal dual algorithm of [39]. A geometric progression of regularization parameters was used for MHDM: $\lambda_n = \lambda_0 10^n$ with $\lambda_0 = 1$. To compare the two methods, the difference between the MHDM iterate x_n and the Tikhonov minimizer x_{λ_n} at scale λ_n was computed via $e_n = \|x_n - x_{\lambda_n}\|_{L^2}$. The numerical results are presented in table 1.

We note that the relative error in the denoising case is of magnitude of at most 10^{-3}, while in the deblurring case it is of order 10^{-1} for $n \geqslant 3$. Up to numerical inaccuracies, this confirms the result from theorem 4.19, suggesting that Tikhonov regularization and the MHDM iteration disagree for the deconvolution problem. Furthermore, the numerical results indicate that the MHDM iterates do not converge to the true solution $x^\dagger$ in the deblurring case. This is demonstrated in figure 2 which displays x_n, as well as the one-step regularized solution with parameter λ_n for $n = 4$ and $n = 11$.

5.2. ℓ^p*-regularization for* $p \in (0,1]$

In the regularization of sparsity constrained ill-posed problems, one often employs Tikhonov regularization with ℓ^1-penalty. Another approach, even more sparsity promoting, is to use the ℓ^p-quasi-norms with $p \in (0,1)$—see, e.g. [40, 41].

[3] The program code is available as ancillary file from the arXiv page of this paper.

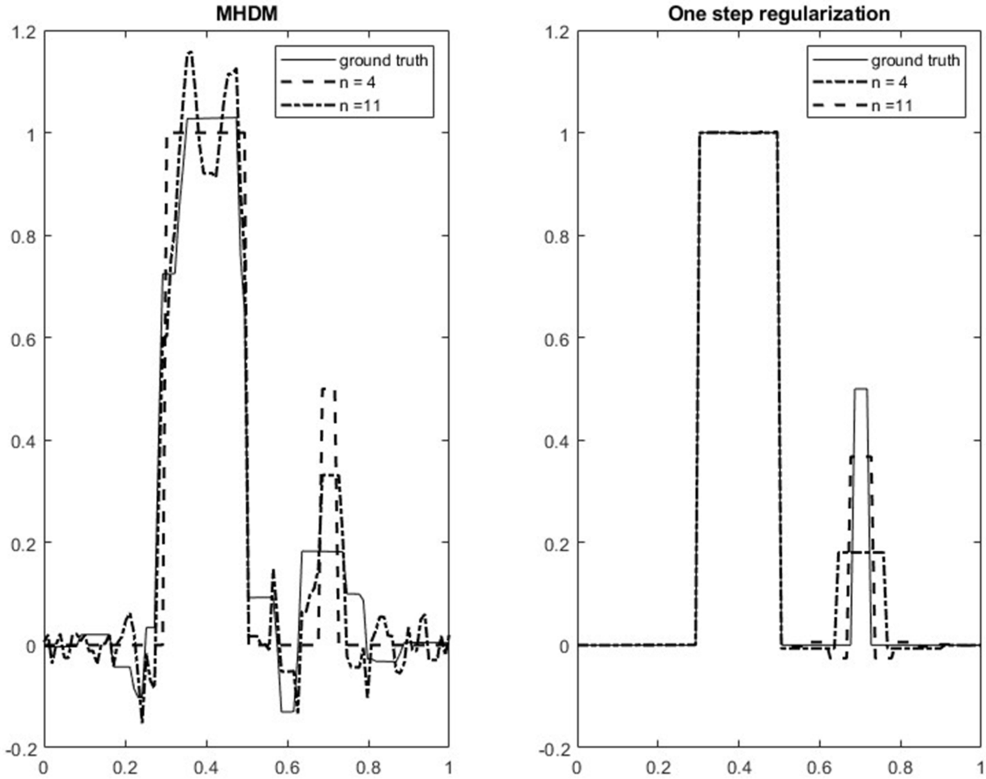

Figure 2. Comparison of different iterates of the MHDM and the corresponding one-step regularizers for TV-deblurring.

In our experiments, we consider a sparse signal with peaks of different amplitudes. We apply a Gaussian convolution operator with standard distribution 0.025 and add a normally distributed noise to create noisy data f^δ (cf figure 3). We use the discrepancy principle for both the MHDM and the Tikhonov regularization. This means that we stop the iteration according to (2.13) for the MHDM, while for solving the Tikhonov regularization problem we consider the same sequence of parameters and stop when the corresponding discrepancy principle condition is satisfied. We choose $\tau = 1.01$ in all experiments.

In general, we do not expect the results of the MHDM to be significantly superior to those of the Tikhonov regularization. This can also be seen in figures 4 and 5. Hence, we are more interested in how robust the algorithm is with regard to the involved parameters.

Let us start with the case of the ℓ^1-penalty. All minimizers were computed using Nesterov's algorithm [42]. In table 2 one can see the relative ℓ^2-errors (that is $\left\|\tilde{x} - x^\dagger\right\|_{\ell^2} \left\|x^\dagger\right\|_{\ell^2}^{-1}$ if $\tilde{x}$ is an approximate solution obtained by either the MHDM or Tikhonov regularization) at the stopping index for different noise levels. For both MHDM and Tikhonov regularization, we use a geometric progression $\lambda_n = 2^n \lambda_0$ with $\lambda_0 = 1$. Note that the Tikhonov regularization did not meet the discrepancy principle in the case of the smallest noise level for any of the first 100 tested parameters (see the $*$-entries in table 2).

While the number of minimizations used to meet the discrepancy principle is comparable, the MHDM performs slightly better than the Tikhonov regularization. Let us now investigate

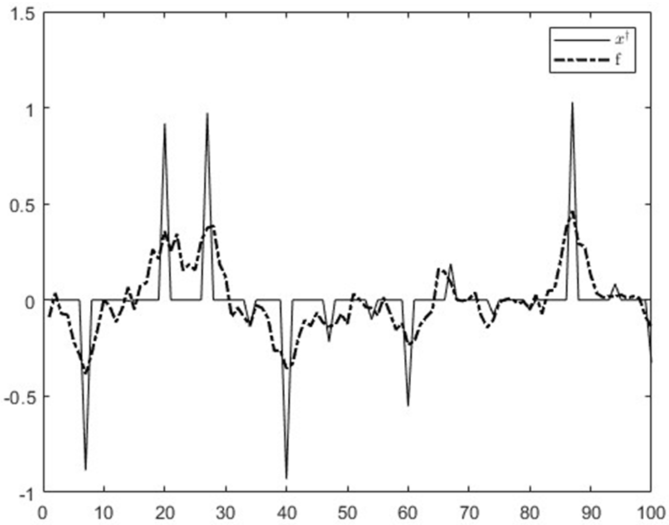

Figure 3. Ground truth $x^\dagger$ and observed data f^δ.

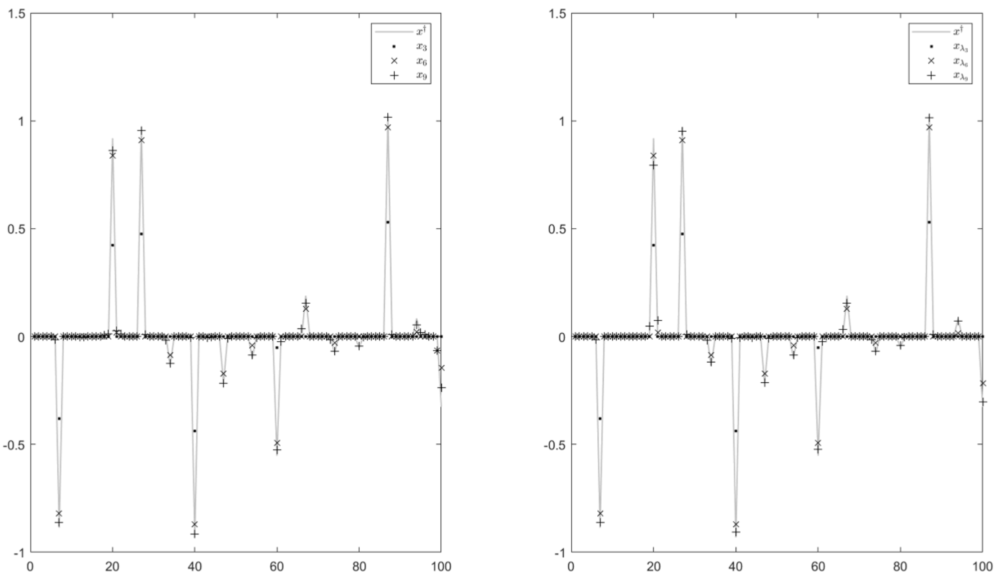

Figure 4. Iterates of the MHDM (left) and corresponding Tikhonov regularizers (right) with the parameters $\lambda_0 = 1$ and $\lambda_n = 2\lambda_{n-1}$.

the stability of the algorithms with respect to the parameter choice. In tables 3 and 4 one can see the relative ℓ^2-errors and stopping indices for varying initial values λ_0 and varying ratios for the geometric progression, respectively.

We observe that, in case $\lambda_n = 2^n\lambda_0$, the choice of the initial guess λ_0 is not too important. However, varying the ratio $\frac{\lambda_{n+1}}{\lambda_n}$ leads to quite different results for Tikhonov regularization, while the MHDM behavior does not change significantly. Therefore, we argue that in the case

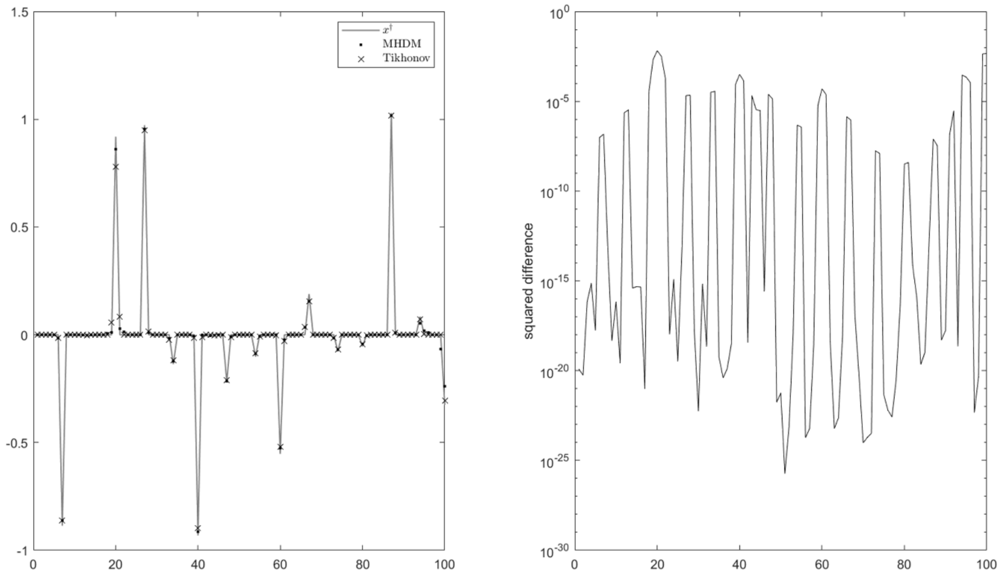

Figure 5. Comparison of the MHDM iterate and Tikhonov regularizer at the respective stopping index (left) and componentwise squared error (right) with the parameters $\lambda_0 = 1$ and $\lambda_n = 2\lambda_{n-1}$.

Table 2. Relative errors under different levels of noise for the ℓ^1-penalty.

δ	0.0051	0.0508	0.5083
Relative error MHDM	0.0585	0.0697	0.4457
Relative error Tikhonov	*	0.0888	0.6040
n^*_{MHDM}	15	9	6
n^*_{Tikhonov}	*	9	6

Table 3. Relative errors for different initial guesses λ_0 with $\lambda_n = 2^n\lambda_0$ and $\delta = 0.0508$ for the ℓ^1-penalty.

λ_0	0.01	0.1	1	10
Relative error MHDM	0.0527	0.0720	0.0697	0.0536
Relative error Tikhonov	0.0981	0.0803	0.0888	0.0973
n^*_{MHDM}	16	13	9	6
n^*_{Tikhonov}	16	12	9	6

of ℓ^1-deblurring, the MHDM is a rather robust method, which in average seems to outperform Tikhonov regularization.

Let us now turn to ℓ^p-regularization for $p \in (0,1)$, that is, consider the functional

$$J_p(x) = \sum_{i=0}^{\infty} |x_i|^p. \tag{5.1}$$

Table 4. Relative errors for different geometric progressions with initial guess $\lambda_0 = 1$ and $\delta = 0.0508$ for the ℓ^1-penalty.

ratio	1.2	2	3	10
Relative error MHDM	0.0717	0.0697	0.0473	0.0594
Relative error Tikhonov	0.0789	0.0803	0.0888	0.1244
n^*_{MHDM}	31	9	6	4
n^*_{Tikhonov}	30	9	6	4

Table 5. Relative errors for different choices of p.

p	0.995	0.9	0.75	0.5	0.25	0.05
Relative error MHDM	0.0512	0.0492	0.0395	0.0594	0.0331	0.0494
Relative error Tikhonov	0.0890	0.0211	0.0132	0.0108	0.0109	0.0108
n^*_{MHDM}	16	16	16	19	20	16
n^*_{Tikhonov}	16	16	17	17	17	18

Table 6. Relative errors for MHDM with varying penalty terms J_{p_n}.

	p_n increasing	p_n decreasing
Relative error	0.0489	0.0533
n^*	15	18
p_{n^*}	0.1500	0.8643

In order to compute minimizers of the generalized Tikhonov functional with this penalty term, we use the algorithm introduced in [43], whose theorem 1 also ensures the well-definedness of the single step regularization and of the MHDM. Note that we may apply the same stopping rule (2.13), since the assumptions of part (*i*) in theorem 2.1 are satisfied with $C = 1$. Indeed, for any $p \in (0,1)$ and $x, y \geqslant 0$, one has

$$|x-y|^p \leqslant |x|^p + |y|^p \,.$$

We first compare the Tikhonov method with the MHDM for $\lambda_n = \lambda_0 2^n$ with $\lambda_0 = 0.01$ and noise level $\delta = 0.0508$, while allowing different values for p. Furthermore, we consider a version of the flexible MHDM (3.1) employing J_n as in (5.1) with a variable p_n instead of a fixed p in each iteration, namely for an increasing sequence $p_n = 0.95 - \frac{0.9}{n+1}$ and then for a decreasing sequence $p_n = 0.05 + \frac{0.9}{n+1}$. The results of both experiments can be found in tables 5 and 6

Once again, the number of minimizations until the discrepancy principle is satisfied is very similar for both MHDM and Tikhonov regularization. For p close to 1, the MHDM seems to produce slightly better results, while for smaller values of p Tikhonov regularization seems to be superior. Those results achieved by Tikhonov regularization are also the overall best ones. The more general approach with functionals J_{p_n} did not show very different results from the approach with fixed exponent. Nevertheless, a version with adaptive penalty terms would be an interesting concept for further research. For variations of λ_0 and of the ratio $\frac{\lambda_n}{\lambda_{n+1}}$ defining the parameters λ_n in the case of fixed exponent p, we observe that the MHDM performs again very similarly. The Tikhonov regularization performs more stable than in the ℓ^1 case, though it is outperformed by the MHDM for large ratios $\left(\frac{\lambda_{n+1}}{\lambda_n} \approx 60\right)$. We expect that for ill-posed problems with higher degree of ill-posedness than the one we considered, the outperformance will occur for smaller ratios. Thus, we conclude that by applying the discrepancy principle, both methods seem to produce comparable reconstructions of the true data, but the MHDM is less sensitive to parameter choices.

6. Conclusion

We analyze the MHDM involving various convex and nonconvex penalties in a general function space framework and provide sufficient conditions for the convergence of the residual. We also provide a counterexample for which the residual does not converge, while the sufficient conditions are not satisfied either. Then, we extend the MHDM to adaptive regularization functionals, showing an interesting multiscale norm decomposition of the data. This applies in particular to the Bregman iteration method, thus leading to a new result in this respect. Furthermore, we propose a characterization for the generalized Tikhonov regularization at a given scale to agree with the MHDM. We provide a sufficient condition for the agreement in finite dimensional ℓ^1-regularization and use it to prove that the MHDM and Tikhonov regularization are identical for 1-dimensional TV-denoising. Moreover, we test the MHDM for sparsity constrained deconvolution problems and find it to be stable with regard to the involved parameters. Conditions for the convergence of the MHDM iterates, as well as convergence rates remain open questions.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

The authors are grateful to Daria Ghilli (University of Pavia) for providing an initial version of a code for nonconvex sparsity regularization. E Resmerita and T Wolf are supported by the Austrian Science Fund (FWF): DOC 78. The constructive remarks of the referees are appreciated, as they led to improving the presentation of the manuscript.

ORCID iDs

Stefan Kindermann https://orcid.org/0000-0002-3688-5125
Elena Resmerita https://orcid.org/0000-0001-6670-8966
Tobias Wolf https://orcid.org/0000-0002-2665-4219

References

[1] Tadmor E, Nezzar S and Vese L 2004 A multiscale image representation using hierarchical (BV,L^2) decompositions *Multiscale Model. Simul.* **2** 554–79
[2] Nezzar S, Tadmor E and Vese L 2008 Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation *Commun. Math. Sci.* **6** 281–307
[3] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[4] Tadmor E and Tan C 2012 Hierarchical construction of bounded solutions of div u = f in critical regularity spaces *Nonlinear Partial Differential Equations: The Abel Symp. 2010* pp 255–69

[5] Modin K, Nachman A and Rondi L 2019 A multiscale theory for image registration and nonlinear inverse problems *Adv. Math.* **346** 1009–66
[6] Han H, Wang Z and Zhang Y 2021 Multiscale approach for two-dimensional diffeomorphic image registration *Multiscale Model. Simul.* **19** 1538–72
[7] Hidane M, Lézoray O, Ta V T and Elmoataz A 2010 Nonlocal multiscale hierarchical decomposition on graphs *Computer Vision—ECCV 2010* pp 638–50
[8] Zhong M 2016 Hierarchical reconstruction method for solving ill-posed linear inverse problems *PhD Thesis* University of Maryland
[9] Hofmann B, Kaltenbacher B, Pöschl C and Scherzer O 2007 A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators *Inverse Problems* **23** 987–1010
[10] Groetsch C 2007 *Stable Approximate Evaluation of Unbounded Operators* vol 1894 (Springer)
[11] Hanke M and Groetsch C W 1998 Nonstationary iterated Tikhonov regularization *J. Optim. Theory Appl.* **98** 37–53
[12] Scherzer O and Groetsch C 2001 Inverse scale space theory for inverse problems *Proc. 3rd Int. Conf. on Scale-Space and Morphology in Computer Vision* vol 2106 pp 317–25
[13] Frick K and Scherzer O 2010 Regularization of ill-posed linear equations by the non-stationary augmented Lagrangian method *J. Integral Equ. Appl.* **22** 217–57
[14] Osher S, Burger M, Goldfarb D, Xu J and Yin W 2005 An iterative regularization method for total variation-based image restoration *Multiscale Model. Simul.* **4** 460–89
[15] Li W, Resmerita E and Vese L A 2021 Multiscale hierarchical image decomposition and refinements: qualitative and quantitative results *SIAM J. Imaging Sci.* **14** 844–77
[16] Efron B, Hastie T, Johnstone I and Tibshirani R 2004 Least angle regression *Ann. Stat.* **32** 407–99
[17] Ekeland I and Témam R 1999 *Convex Analysis and Variational Problems* (Society for Industrial and Applied Mathematics)
[18] Lorenz D and Resmerita E 2017 Flexible sparse regularization *Inverse Problems* **33** 014002
[19] Amato U and Hughes W 1991 Maximum-entropy regularization of Fredholm inegral-equations of the first kind *Inverse Problems* **7** 793–808
[20] Engl H W and Landl G 1993 Convergence rates for maximum entropy regularization *SIAM J. Numer. Anal.* **30** 1509–36
[21] Resmerita E and Anderssen R S 2007 Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems *Math. Methods Appl. Sci.* **30** 1527–44
[22] Clason C, Kaltenbacher B and Resmerita E 2019 Regularization of ill-posed problems with non-negative solutions *Splitting Algorithms, Modern Operator Theory and Applications* (Springer) pp 113–35
[23] Combettes P L and Pesquet J C 2011 Proximal splitting methods in signal processing *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (Springer) pp 185–212
[24] Bauschke H H and Combettes P L 2011 *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (Springer)
[25] Zalinescu C 2002 *Convex Analysis in General Vector Spaces* (World Scientific)
[26] Meyer Y 2001 *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations* (American Mathematical Society)
[27] Vese L A and Guyader C L 2016 *Variational Methods in Image Processing* (Chapman & Hall/CRC)
[28] Andreu-Vaillo F, Mazon F, Caselles V and Mazón J 2004 *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals* (*Progress in Mathematics*) (Birkhäuser)
[29] Duan J, Soussen C, Brie D, Idier J and Wang Y P 2011 A sufficient condition on monotonic increase of the number of nonzero entry in the optimizer of L^1 norm penalized least-square problem (arXiv:1104.3792)
[30] Meinshausen N 2007 Relaxed lasso *Comput. Stat. Data Anal.* **52** 374–93
[31] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (Springer)
[32] Pöschl C, Resmerita E and Scherzer O 2010 Discretization of variational regularization in Banach spaces *Inverse Problems* **26** 105017
[33] Chambolle A, Caselles V and Alter F 2005 Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow *Interfaces Free Boundaries* **7** 29–53
[34] Alter F, Caselles V and Chambolle A 2005 A characterization of convex calibrable sets in $\mathbb{R}^n$ *Math. Ann.* **332** 329–66

[35] Allard W K 2008 Total variation regularization for image denoising. I. Geometric theory *SIAM J. Math. Anal.* **39** 1150–90
[36] Condat L 2017 Discrete total variation: new definition and minimization *SIAM J. Imaging Sci.* **10** 1258–90
[37] Ambrosio L, Fusco N and Pallara D 2000 *Functions of Bounded Variation and Free Discontinuity Problems* (*Oxford Mathematical Monographs*) (Oxford University Press)
[38] Bellettini G, Caselles V and Novaga M 2002 The total variation flow in R^N *J. Differ. Equ.* **184** 475–525
[39] Chambolle A and Pock T 2011 A first-order primal-dual algorithm for convex problems with applications to imaging *J. Math. Imaging Vis.* **40** 120–45
[40] Grasmair M 2009 Well-posedness and convergence rates for sparse regularization with sublinear l^q penalty term *Inverse Problems Imaging* **3** 383–7
[41] Lorenz D A 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints *J. Inverse Ill-Posed Probl.* **16** 463–78
[42] Nesterov Y 1983 A method for solving the convex programming problem with convergence rate $o(1/k^2)$ *Proc. USSR Acad. Sci.* **269** 543–7
[43] Ghilli D and Kunisch K 2019 On the monotone and primal-dual active set schemes for ℓ^p-type problems, $p \in (0,1]$ *Comput. Optim. Appl.* **72** 45–85