

PROXIMAL THRESHOLDING ALGORITHM FOR MINIMIZATION OVER ORTHONORMAL BASES

Patrick L. Combettes¹ and Jean-Christophe Pesquet²

¹Laboratoire Jacques-Louis Lions – UMR CNRS 7598
Université Pierre et Marie Curie – Paris 6
75005 Paris, France
plc@math.jussieu.fr

²Institut Gaspard Monge and UMR CNRS 8049
Université de Marne la Vallée
77454 Marne la Vallée Cedex 2, France
pesquet@univ-mlv.fr

September 9, 2006

Abstract

The notion of soft thresholding plays a central role in problems from various areas of applied mathematics, in which the ideal solution is known to possess a sparse decomposition in some orthonormal basis. Using convex-analytical tools, we extend this notion to that of proximal thresholding and investigate its properties, providing in particular several characterizations of such thresholders. We then propose a versatile convex variational formulation for optimization over orthonormal bases that covers a wide range of problems, and establish the strong convergence of a proximal thresholding algorithm to solve it. Numerical applications to signal recovery are demonstrated.

1 Problem formulation

Throughout this paper, \mathcal{H} is a separable infinite-dimensional real Hilbert space with scalar product $\langle \cdot | \cdot \rangle$, norm $\|\cdot\|$, and distance d . Moreover, $\Gamma_0(\mathcal{H})$ denotes the class of proper lower semicontinuous convex functions from \mathcal{H} to $]-\infty, +\infty]$, and $(e_k)_{k \in \mathbb{N}}$ is an orthonormal basis of \mathcal{H} .

The standard denoising problem in signal theory consists of recovering the original form of a signal $\bar{x} \in \mathcal{H}$ from an observation $z = \bar{x} + v$, where $v \in \mathcal{H}$ is the realization of a noise process. In many instances, \bar{x} is known to admit a sparse representation with respect to $(e_k)_{k \in \mathbb{N}}$ and an estimate x of \bar{x} can be constructed by removing the coefficients of smallest magnitude in the

representation $(\langle z | e_k \rangle)_{k \in \mathbb{N}}$ of z with respect to $(e_k)_{k \in \mathbb{N}}$. A popular method consists of performing a so-called soft thresholding of each coefficient $\langle z | e_k \rangle$ at some predetermined level $\omega_k \in]0, +\infty[$, namely (see Fig. 1)

$$(1.1) \quad x = \sum_{k \in \mathbb{N}} \text{soft}_{[-\omega_k, \omega_k]}(\langle z | e_k \rangle) e_k, \quad \text{where} \quad \text{soft}_{[-\omega_k, \omega_k]} : \xi \mapsto \text{sign}(\xi) \max\{|\xi| - \omega_k, 0\}.$$

This approach has received considerable attention in various areas of applied mathematics ranging from nonlinear approximation theory to statistics, and from harmonic analysis to image processing; see for instance [2, 7, 8, 19, 21, 27, 31] and the references therein. From an optimization point of view, the vector x exhibited in (1.1) is simply the solution to the variational problem

$$(1.2) \quad \underset{x \in \mathcal{H}}{\text{minimize}} \quad \frac{1}{2} \|x - z\|^2 + \sum_{k \in \mathbb{N}} \omega_k |\langle x | e_k \rangle|.$$

Attempts have been made to extend this formulation to the more general inverse problems in which the observation assumes the form $z = T\bar{x} + v$, where T is a nonzero bounded linear operator from \mathcal{H} to some real Hilbert space \mathcal{G} , and where $v \in \mathcal{G}$ is the realization of a noise process. Thus, the variational problem

$$(1.3) \quad \underset{x \in \mathcal{H}}{\text{minimize}} \quad \frac{1}{2} \|Tx - z\|^2 + \sum_{k \in \mathbb{N}} \omega_k |\langle x | e_k \rangle|$$

was considered in [5, 17, 18, 22] (see also [34] and the references therein for related work), and the soft thresholding algorithm

$$(1.4) \quad x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \sum_{k \in \mathbb{N}} \text{soft}_{[-\omega_k, \omega_k]}(\langle x_n + T^*(z - Tx_n) | e_k \rangle) e_k$$

was proposed to solve it. The strong convergence of this algorithm was formally established in [16].

Proposition 1.1 [16, Theorem 3.1] *Suppose that $\inf_{k \in \mathbb{N}} \omega_k > 0$ and that $\|T\| < 1$. Then the sequence $(x_n)_{n \in \mathbb{N}}$ generated by (1.4) converges strongly to a solution to (1.3).*

In [14], (1.3) was analyzed in a broader framework and the following extension of Proposition 1.1 was obtained by bringing into play tools from convex analysis and recent results from constructive fixed point theory.

Proposition 1.2 [14, Corollary 5.19] *Let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $]0, 1[$. Suppose that the following hold: $\inf_{k \in \mathbb{N}} \omega_k > 0$, $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} \gamma_n < 2/\|T\|^2$, and $\inf_{n \in \mathbb{N}} \lambda_n > 0$. Then the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the algorithm*

$$(1.5) \quad x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \left(\sum_{k \in \mathbb{N}} \text{soft}_{[-\gamma_n \omega_k, \gamma_n \omega_k]}(\langle x_n + \gamma_n T^*(z - Tx_n) | e_k \rangle) e_k - x_n \right)$$

converges strongly to a solution to (1.3).

In denoising and approximation problems, various theoretical, physical, and heuristic considerations have led researchers to consider alternative thresholding strategies in (1.1); see, e.g., [1, 31, 32, 33, 37]. The same motivations naturally serve as a thrust to investigate the use of alternative thresholding rules in more general algorithms such as (1.5), and to identify the underlying variational problems. These questions are significant because the current theory of iterative thresholding, as described by Proposition 1.2, can tackle only problems described by the variational formulation (1.3), which offers limited flexibility in the penalization of the coefficients $(\langle x | e_k \rangle)_{k \in \mathbb{N}}$ and which is furthermore restricted to standard linear inverse problems. The aim of the present paper is to bring out general answers to these questions. Our analysis will revolve around the following variational formulation, where σ_Ω denotes the support function of a set Ω .

Problem 1.3 Let $\Phi \in \Gamma_0(\mathcal{H})$, let $\mathbb{K} \subset \mathbb{N}$, let $\mathbb{L} = \mathbb{N} \setminus \mathbb{K}$, let $(\Omega_k)_{k \in \mathbb{K}}$ be a sequence of closed intervals in \mathbb{R} , and let $(\psi_k)_{k \in \mathbb{N}}$ be a sequence in $\Gamma_0(\mathbb{R})$. The objective is to

$$(1.6) \quad \underset{x \in \mathcal{H}}{\text{minimize}} \quad \Phi(x) + \sum_{k \in \mathbb{N}} \psi_k(\langle x | e_k \rangle) + \sum_{k \in \mathbb{K}} \sigma_{\Omega_k}(\langle x | e_k \rangle),$$

under the following standing assumptions:

- (i) the function Φ is differentiable on \mathcal{H} , $\inf \Phi(\mathcal{H}) > -\infty$, and $\nabla \Phi$ is $1/\beta$ -Lipschitz continuous for some $\beta \in]0, +\infty[$;
- (ii) for every $k \in \mathbb{N}$, $\psi_k \geq \psi_k(0) = 0$;
- (iii) the functions $(\psi_k)_{k \in \mathbb{N}}$ are differentiable at 0;
- (iv) the functions $(\psi_k)_{k \in \mathbb{L}}$ are finite and twice differentiable on $\mathbb{R} \setminus \{0\}$, and
$$(1.7) \quad (\forall \rho \in]0, +\infty[)(\exists \theta \in]0, +\infty[) \quad \inf_{k \in \mathbb{L}} \inf_{0 < |\xi| \leq \rho} \psi_k''(\xi) \geq \theta;$$
- (v) the function $\Upsilon_{\mathbb{L}}: \ell^2(\mathbb{L}) \rightarrow]-\infty, +\infty]: (\xi_k)_{k \in \mathbb{L}} \mapsto \sum_{k \in \mathbb{L}} \psi_k(\xi_k)$ is coercive;
- (vi) $0 \in \text{int} \bigcap_{k \in \mathbb{K}} \Omega_k$.

Let us note that Problem 1.3 reduces to (1.3) when $\Phi: x \mapsto \|Tx - z\|^2/2$, $\mathbb{K} = \mathbb{N}$, and, for every $k \in \mathbb{N}$, $\Omega_k = [-\omega_k, \omega_k]$ and $\psi_k = 0$. It will be shown (Proposition 4.1) that Problem 1.3 admits at least one solution. While assumption (i) on Φ may seem offhand to be rather restrictive, it will be seen in Section 5.1 to cover important scenarios. In addition, it makes it possible to employ a forward-backward splitting strategy to solve (1.6), which consists essentially of alternating a forward (explicit) gradient step on Φ with a backward (implicit) proximal step on

$$(1.8) \quad \Psi: \mathcal{H} \rightarrow]-\infty, +\infty]: x \mapsto \sum_{k \in \mathbb{N}} \psi_k(\langle x | e_k \rangle) + \sum_{k \in \mathbb{K}} \sigma_{\Omega_k}(\langle x | e_k \rangle).$$

Our main convergence result (Theorem 4.5) will establish the *strong* convergence of an inexact forward-backward splitting algorithm (Algorithm 4.3) for solving Problem 1.3. Another contribution of this paper will be to show (Remark 3.4) that, under our standing assumptions, the function

displayed in (1.8) is quite general in the sense that the operators on \mathcal{H} that perform nonexpansive (as required by our convergence analysis) and nonincreasing (as imposed by practical considerations) thresholdings on the closed intervals $(\Omega_k)_{k \in \mathbb{K}}$ of the coefficients $(\langle x | e_k \rangle)_{k \in \mathbb{K}}$ of a point $x \in \mathcal{H}$ are precisely those of the form prox_Ψ , i.e., the proximity operator of Ψ . Furthermore, we show (Proposition 3.5 and Lemma 2.3) that such an operator, which provides the proximal step of our algorithm, can be conveniently decomposed as

$$(1.9) \quad \text{prox}_\Psi: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \sum_{k \in \mathbb{K}} \text{prox}_{\psi_k}(\text{soft}_{\Omega_k} \langle x | e_k \rangle) e_k + \sum_{k \in \mathbb{L}} \text{prox}_{\psi_k} \langle x | e_k \rangle e_k,$$

where we define the soft thresholder relative to a nonempty closed interval $\Omega \subset \mathbb{R}$ as

$$(1.10) \quad \text{soft}_\Omega: \mathbb{R} \rightarrow \mathbb{R}: \xi \mapsto \begin{cases} \xi - \underline{\omega}, & \text{if } \xi < \underline{\omega}; \\ 0, & \text{if } \xi \in \Omega; \\ \xi - \bar{\omega}, & \text{if } \xi > \bar{\omega}, \end{cases} \quad \text{with} \quad \begin{cases} \underline{\omega} = \inf \Omega, \\ \bar{\omega} = \sup \Omega. \end{cases}$$

The remainder of the paper is organized as follows. In Section 2, we provide a brief account of the theory of proximity operators, which play a central role in our analysis. In Section 3, we introduce and study the notion of a proximal thresholder. Our algorithm is presented in Section 4 and its strong convergence to a solution to Problem 1.3 is demonstrated. Signal recovery applications are discussed in Section 5, where numerical results are presented.

2 Proximity operators

Let us first introduce some basic notation (for a detailed account of convex analysis, see [39]). Let C be a subset of \mathcal{H} . The indicator function of C is

$$(2.1) \quad \iota_C: \mathcal{H} \rightarrow \{0, +\infty\}: x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C, \end{cases}$$

its support function is $\sigma_C: \mathcal{H} \rightarrow [-\infty, +\infty]: u \mapsto \sup_{x \in C} \langle x | u \rangle$, and its distance function is $d_C: \mathcal{H} \rightarrow [0, +\infty]: x \mapsto \inf \|C - x\|$. If C is nonempty, closed, and convex then, for every $x \in \mathcal{H}$, there exists a unique point $P_C x \in C$, called the projection of x onto C , such that $\|x - P_C x\| = d_C(x)$. A function $f: \mathcal{H} \rightarrow [-\infty, +\infty]$ is proper if $-\infty \notin f(\mathcal{H}) \neq \{+\infty\}$. The domain of $f: \mathcal{H} \rightarrow [-\infty, +\infty]$ is $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$, its set of global minimizers is denoted by $\text{Argmin } f$, and its conjugate is the function $f^*: \mathcal{H} \rightarrow [-\infty, +\infty]: u \mapsto \sup_{x \in \mathcal{H}} \langle x | u \rangle - f(x)$; if f is proper, its subdifferential is the set-valued operator

$$(2.2) \quad \partial f: \mathcal{H} \rightarrow 2^{\mathcal{H}}: x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \text{dom } f) \langle y - x | u \rangle + f(x) \leq f(y)\}.$$

If $f: \mathcal{H} \rightarrow]-\infty, +\infty]$ is convex and Gâteaux differentiable at $x \in \text{dom } f$ with gradient $\nabla f(x)$, then $\partial f(x) = \{\nabla f(x)\}$.

Example 2.1 *Let $\Omega \subset \mathbb{R}$ be a nonempty closed interval, let $\underline{\omega} = \inf \Omega$, let $\bar{\omega} = \sup \Omega$, and let $\xi \in \mathbb{R}$. Then the following hold.*

$$(i) \quad \sigma_{\Omega}(\xi) = \begin{cases} \underline{\omega}\xi, & \text{if } \xi < 0; \\ 0, & \text{if } \xi = 0; \\ \overline{\omega}\xi, & \text{if } \xi > 0. \end{cases}$$

$$(ii) \quad \partial\sigma_{\Omega}(\xi) = \begin{cases} \{\underline{\omega}\} \cap \mathbb{R}, & \text{if } \xi < 0; \\ \Omega, & \text{if } \xi = 0; \\ \{\overline{\omega}\} \cap \mathbb{R}, & \text{if } \xi > 0. \end{cases}$$

The infimal convolution of two functions $f, g: \mathcal{H} \rightarrow]-\infty, +\infty]$ is denoted by $f \square g$. Finally, an operator $T: \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive if $(\forall(x, y) \in \mathcal{H}^2) \|Tx - Ty\| \leq \|x - y\|$ and firmly nonexpansive if $(\forall(x, y) \in \mathcal{H}^2) \|Tx - Ty\|^2 \leq \langle x - y \mid Tx - Ty \rangle$.

Proximity operators were introduced by Moreau [28]. We briefly recall some essential facts below and refer the reader to [14] and [29] for more details. Let $f \in \Gamma_0(\mathcal{H})$. The proximity operator of f is the operator $\text{prox}_f: \mathcal{H} \rightarrow \mathcal{H}$ which maps every $x \in \mathcal{H}$ to the unique minimizer of the function $y \mapsto f(y) + \|x - y\|^2/2$. It is characterized by

$$(2.3) \quad (\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \text{prox}_f x \quad \Leftrightarrow \quad x - p \in \partial f(p).$$

Lemma 2.2 *Let $f \in \Gamma_0(\mathcal{H})$. Then the following hold.*

$$(i) \quad (\forall x \in \mathcal{H}) \quad x \in \text{Argmin } f \Leftrightarrow 0 \in \partial f(x) \Leftrightarrow \text{prox}_f x = x.$$

$$(ii) \quad \text{prox}_{f^*} = \text{Id} - \text{prox}_f.$$

(iii) prox_f is firmly nonexpansive.

(iv) If f is even, then prox_f is odd.

Lemma 2.3 [14, Example 2.19] *Let $(b_k)_{k \in \mathbb{N}}$ be an orthonormal basis of \mathcal{H} and let*

$$(2.4) \quad f: \mathcal{H} \rightarrow]-\infty, +\infty] : x \mapsto \sum_{k \in \mathbb{N}} \phi_k(\langle x \mid b_k \rangle),$$

where $(\phi_k)_{k \in \mathbb{N}}$ are functions in $\Gamma_0(\mathbb{R})$ that satisfy $(\forall k \in \mathbb{N}) \phi_k \geq \phi_k(0) = 0$. Then $f \in \Gamma_0(\mathcal{H})$ and $(\forall x \in \mathcal{H}) \text{prox}_f x = \sum_{k \in \mathbb{N}} \text{prox}_{\phi_k} \langle x \mid b_k \rangle b_k$.

The remainder of this section is dedicated to proximity operators on the real line, the importance of which is underscored by Lemma 2.3.

Proposition 2.4 *Let ϱ be a function defined from \mathbb{R} to \mathbb{R} . Then ϱ is the proximity operator of a function in $\Gamma_0(\mathbb{R})$ if and only if it is nonexpansive and nondecreasing.*

Proof. Let ξ and η be real numbers. First, suppose that $\varrho = \text{prox}_{\phi}$, where $\phi \in \Gamma_0(\mathbb{R})$. Then it follows from Lemma 2.2(iii) that ϱ is nonexpansive and that $0 \leq |\varrho(\xi) - \varrho(\eta)|^2 \leq (\xi - \eta)(\varrho(\xi) - \varrho(\eta))$, which shows that ϱ is nondecreasing since $\xi - \eta$ and $\varrho(\xi) - \varrho(\eta)$ have the same sign. Conversely,

suppose that ϱ is nonexpansive and nondecreasing and, without loss of generality, that $\xi \leq \eta$. Then, $0 \leq \varrho(\xi) - \varrho(\eta) \leq \xi - \eta$ and therefore $|\varrho(\xi) - \varrho(\eta)|^2 \leq (\xi - \eta)(\varrho(\xi) - \varrho(\eta))$. Thus, ϱ is firmly nonexpansive. However, every firmly nonexpansive operator $T: \mathcal{H} \rightarrow \mathcal{H}$ is of the form $T = (\text{Id} + A)^{-1}$, where $A: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ is a maximal monotone operator [6]. Since the only maximal monotone operators in \mathbb{R} are subdifferentials of functions in $\Gamma_0(\mathbb{R})$ [30, Section 24], we must have $\varrho = (\text{Id} + \partial\phi)^{-1} = \text{prox}_\phi$ for some $\phi \in \Gamma_0(\mathbb{R})$. \square

Corollary 2.5 *Suppose that $\phi \in \Gamma_0(\mathbb{R})$ is minimized by 0. Then*

$$(2.5) \quad (\forall \xi \in \mathbb{R}) \quad \begin{cases} 0 \leq \text{prox}_\phi \xi \leq \xi, & \text{if } \xi > 0; \\ \text{prox}_\phi \xi = 0, & \text{if } \xi = 0; \\ \xi \leq \text{prox}_\phi \xi \leq 0, & \text{if } \xi < 0. \end{cases}$$

This is true in particular when ϕ is even, in which case prox_ϕ is an odd operator.

Proof. Since $0 \in \text{Argmin } \phi$, Lemma 2.2(i) yields $\text{prox}_\phi 0 = 0$. In turn, since prox_ϕ is nonexpansive by Lemma 2.2(iii), we have $(\forall \xi \in \mathbb{R}) |\text{prox}_\phi \xi| = |\text{prox}_\phi \xi - \text{prox}_\phi 0| \leq |\xi - 0| = |\xi|$. Altogether, since Proposition 2.4 asserts that prox_ϕ is nondecreasing, we obtain (2.5). Finally, if ϕ is even, its convexity yields $(\forall \xi \in \text{dom } \phi) \phi(0) = \phi((\xi - \xi)/2) \leq (\phi(\xi) + \phi(-\xi))/2 = \phi(\xi)$. Therefore $0 \in \text{Argmin } \phi$, while the oddness of prox_ϕ follows from Lemma 2.2(iv). \square

Let us now provide some elementary examples (Example 2.6 is illustrated in Fig. 1 in the case when $\Omega = [-1, 1]$).

Example 2.6 *Let $\Omega \subset \mathbb{R}$ be a nonempty closed interval, let $\underline{\omega} = \inf \Omega$, let $\bar{\omega} = \sup \Omega$, and let $\xi \in \mathbb{R}$. Then the following hold.*

$$(i) \quad \text{prox}_{\iota_\Omega} \xi = P_\Omega \xi = \begin{cases} \underline{\omega}, & \text{if } \xi < \underline{\omega}; \\ \xi, & \text{if } \xi \in \Omega; \\ \bar{\omega}, & \text{if } \xi > \bar{\omega}. \end{cases}$$

$$(ii) \quad \text{prox}_{\sigma_\Omega} \xi = \text{soft}_\Omega \xi, \text{ where } \text{soft}_\Omega \text{ is the soft thresholder defined in (1.10).}$$

Proof. (i) is clear and, since $\sigma_\Omega^* = \iota_\Omega$, (ii) follows from (i) and Lemma 2.2(ii). \square

Example 2.7 *Let $p \in [1, +\infty[$, let $\omega \in]0, +\infty[$, let $\phi: \mathbb{R} \rightarrow \mathbb{R}: \eta \mapsto \omega|\eta|^p$, let $\xi \in \mathbb{R}$, and set $\pi = \text{prox}_\phi \xi$. Then the following hold.*

$$(i) \quad \pi = \text{sign}(\xi) \max\{|\xi| - \omega, 0\}, \text{ if } p = 1;$$

$$(ii) \quad \pi = \xi + \frac{4\omega}{3 \cdot 2^{1/3}} \left((\rho - \xi)^{1/3} - (\rho + \xi)^{1/3} \right), \text{ where } \rho = \sqrt{\xi^2 + 256\omega^3/729}, \text{ if } p = 4/3;$$

$$(iii) \quad \pi = \xi + 9\omega^2 \text{sign}(\xi) \left(1 - \sqrt{1 + 16|\xi|/(9\omega^2)} \right) / 8, \text{ if } p = 3/2;$$

$$(iv) \quad \pi = \xi / (1 + 2\omega), \text{ if } p = 2;$$

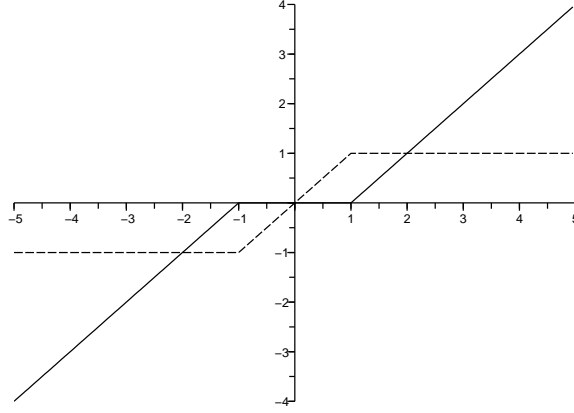


Figure 1: Graphs of $\text{prox}_\phi = \text{soft}_{[-1,1]}$ (solid line) and $\text{prox}_{\phi^*} = P_{[-1,1]}$ (dashed line), where $\phi = |\cdot|$ and $\phi^* = \iota_{[-1,1]}$.

(v) $\pi = \text{sign}(\xi)(\sqrt{1 + 12\omega|\xi|} - 1)/(6\omega)$, if $p = 3$;

(vi) $\pi = \left(\frac{\rho + \xi}{8\omega}\right)^{1/3} - \left(\frac{\rho - \xi}{8\omega}\right)^{1/3}$, where $\rho = \sqrt{\xi^2 + 1/(27\omega)}$, if $p = 4$.

Proof. (i): Set $\Omega = [-\omega, \omega]$ in Example 2.6(ii). (ii)–(vi): Since ϕ is even, we can assume that $\xi \geq 0$ and then extend the result for $\xi \leq 0$ by antisymmetry via Corollary 2.5. As ϕ is differentiable, it follows from (2.3) and Corollary 2.5 that π is the unique nonnegative solution to $\xi - \pi = \phi'(\pi) = p\omega\pi^{p-1}$, which can be solved explicitly in each case. \square

Proposition 2.8 *Let ψ be a function in $\Gamma_0(\mathbb{R})$, and let ρ and θ be real numbers in $]0, +\infty[$ such that:*

- (i) $\psi \geq \psi(0) = 0$;
- (ii) ψ is differentiable at 0;
- (iii) ψ is twice differentiable on $[-\rho, \rho] \setminus \{0\}$ and $\inf_{0 < |\xi| \leq \rho} \psi''(\xi) \geq \theta$.

Then $(\forall \xi \in [-\rho, \rho])(\forall \eta \in [-\rho, \rho]) |\text{prox}_\psi \xi - \text{prox}_\psi \eta| \leq |\xi - \eta|/(1 + \theta)$.

Proof. Set $R = [-\rho, \rho] \setminus \{0\}$ and $\varphi: R \rightarrow \mathbb{R}: \zeta \mapsto \zeta + \psi'(\zeta)$. We first infer from (iii) that

$$(2.6) \quad (\forall \zeta \in R) \quad \varphi'(\zeta) = 1 + \psi''(\zeta) \geq 1 + \theta.$$

Moreover, (2.3) yields $(\forall \zeta \in R) \operatorname{prox}_\psi \zeta = \varphi^{-1}(\zeta)$. Note also that, in the light of (2.3), (ii), and (i), we have $(\forall \zeta \in \mathbb{R}) \operatorname{prox}_\psi \zeta = 0 \Leftrightarrow \zeta \in \partial\psi(0) = \{\psi'(0)\} = \{0\}$. Hence, prox_ψ vanishes only at 0 and we derive from Lemma 2.2(iii) that

$$(2.7) \quad (\forall \zeta \in R) \quad 0 < |\varphi^{-1}(\zeta)| = |\operatorname{prox}_\psi \zeta - \operatorname{prox}_\psi 0| \leq |\zeta - 0| \leq \rho.$$

In turn, we deduce from (2.6) that

$$(2.8) \quad \sup_{\zeta \in R} \operatorname{prox}'_\psi \zeta = \frac{1}{\inf_{\zeta \in R} \varphi'(\varphi^{-1}(\zeta))} \leq \frac{1}{\inf_{\zeta \in R} \varphi'(\zeta)} \leq \frac{1}{1 + \theta}.$$

Now fix ξ and η in R . First, let us assume that either $\xi < \eta < 0$ or $0 < \xi < \eta$. Then, since prox_ψ is nondecreasing by Proposition 2.4, it follows from the mean value theorem and (2.8) that there exists $\mu \in]\xi, \eta[$ such that

$$(2.9) \quad 0 \leq \operatorname{prox}_\psi \eta - \operatorname{prox}_\psi \xi = (\eta - \xi) \operatorname{prox}'_\psi \mu \leq (\eta - \xi) \sup_{\zeta \in R} \operatorname{prox}'_\psi \zeta \leq \frac{\eta - \xi}{1 + \theta}.$$

Next, let us assume that $\xi < 0 < \eta$. Then the mean value theorem asserts that there exist $\mu \in]\xi, 0[$ and $\nu \in]0, \eta[$ such that

$$(2.10) \quad \operatorname{prox}_\psi 0 - \operatorname{prox}_\psi \xi = -\xi \operatorname{prox}'_\psi \mu \quad \text{and} \quad \operatorname{prox}_\psi \eta - \operatorname{prox}_\psi 0 = \eta \operatorname{prox}'_\psi \nu.$$

Since prox_ψ is nondecreasing and $\operatorname{prox}_\psi 0 = 0$, we obtain

$$(2.11) \quad 0 \leq \operatorname{prox}_\psi \eta - \operatorname{prox}_\psi \xi = \eta \operatorname{prox}'_\psi \nu - \xi \operatorname{prox}'_\psi \mu \leq (\eta - \xi) \sup_{\zeta \in R} \operatorname{prox}'_\psi \zeta \leq \frac{\eta - \xi}{1 + \theta}.$$

Altogether, we have shown that, for every ξ and η in R , $|\operatorname{prox}_\psi \xi - \operatorname{prox}_\psi \eta| \leq |\xi - \eta|/(1 + \theta)$. We conclude by observing that, due to the continuity of prox_ψ (Lemma 2.2(iii)), this inequality holds for every ξ and η in $[-\rho, \rho]$. \square

3 Proximal thresholding

The standard soft thresholder of (1.1), which was extended to closed intervals in (1.10), was seen in Example 2.6(ii) to be a proximity operator. As such, it possesses attractive properties (see Lemma 2.2(i)&(iii)) that prove extremely useful in the convergence analysis of iterative methods [11]. This remark motivates the following definition.

Definition 3.1 Let $T: \mathcal{H} \rightarrow \mathcal{H}$ and let Ω be a nonempty closed convex subset of \mathcal{H} . Then T is a *proximal thresholder* on Ω if there exists a function $f \in \Gamma_0(\mathcal{H})$ such that

$$(3.1) \quad T = \operatorname{prox}_f \quad \text{and} \quad (\forall x \in \mathcal{H}) \quad Tx = 0 \Leftrightarrow x \in \Omega.$$

The next proposition provides characterizations of proximal thresholders.

Proposition 3.2 *Let $f \in \Gamma_0(\mathcal{H})$ and let Ω be a nonempty closed convex subset of \mathcal{H} . Then the following are equivalent.*

- (i) prox_f is a proximal thresholder on Ω .
- (ii) $\partial f(0) = \Omega$.
- (iii) $(\forall x \in \mathcal{H}) [\text{prox}_{f^*} x = x \Leftrightarrow x \in \Omega]$.
- (iv) $\text{Argmin } f^* = \Omega$.

In particular, (i)–(iv) hold when

- (v) $f = g + \sigma_\Omega$, where $g \in \Gamma_0(\mathcal{H})$ is Gâteaux differentiable at 0 and $\nabla g(0) = 0$.

Proof. (i) \Leftrightarrow (ii): Fix $x \in \mathcal{H}$. Then it follows from (2.3) that

$$(3.2) \quad \begin{aligned} [\text{prox}_f x = 0 \Leftrightarrow x \in \Omega] &\Leftrightarrow [x \in \partial f(0) \Leftrightarrow x \in \Omega] \\ &\Leftrightarrow \partial f(0) = \Omega. \end{aligned}$$

(i) \Leftrightarrow (iii): Fix $x \in \mathcal{H}$. Then it follows from Lemma 2.2(ii) that

$$(3.3) \quad [\text{prox}_f x = 0 \Leftrightarrow x \in \Omega] \Leftrightarrow [x - \text{prox}_{f^*} x = 0 \Leftrightarrow x \in \Omega].$$

(iii) \Leftrightarrow (iv): Since $f \in \Gamma_0(\mathcal{H})$, $f^* \in \Gamma_0(\mathcal{H})$ and we can apply Lemma 2.2(i) to f^* .

(v) \Rightarrow (ii): Since (v) implies that $0 \in \text{core dom } g$, we have $0 \in (\text{core dom } g) \cap \text{dom } \sigma_\Omega$ and it follows from [39, Theorem 2.8.3] that

$$(3.4) \quad \partial f(0) = \partial(g + \sigma_\Omega)(0) = \partial g(0) + \partial \sigma_\Omega(0) = \partial g(0) + \Omega,$$

where the last equality results from the observation that, for every $u \in \mathcal{H}$, Fenchel's identity yields $u \in \partial \sigma_\Omega(0) \Leftrightarrow 0 = \langle 0 | u \rangle = \sigma_\Omega(0) + \sigma_\Omega^*(u) \Leftrightarrow 0 = \sigma_\Omega^*(u) = \iota_\Omega(u) \Leftrightarrow u \in \Omega$. However, since $\partial g(0) = \{\nabla g(0)\} = \{0\}$, we obtain $\partial f(0) = \Omega$, and (ii) is therefore satisfied. \square

The following theorem is a significant refinement of a result of Proposition 3.2 in the case when $\mathcal{H} = \mathbb{R}$, that characterizes all the functions $\phi \in \Gamma_0(\mathbb{R})$ for which prox_ϕ is a proximal thresholder.

Theorem 3.3 *Let $\phi \in \Gamma_0(\mathbb{R})$ and let $\Omega \subset \mathbb{R}$ be a nonempty closed interval. Then the following are equivalent.*

- (i) prox_ϕ is a proximal thresholder on Ω .
- (ii) $\phi = \psi + \sigma_\Omega$, where $\psi \in \Gamma_0(\mathbb{R})$ is differentiable at 0 and $\psi'(0) = 0$.

Proof. In view of Proposition 3.2, it is enough to show that $\partial\phi(0) = \Omega \Rightarrow$ (ii). So let us assume that $\partial\phi(0) = \Omega$, and set $\underline{\omega} = \inf \Omega$ and $\bar{\omega} = \sup \Omega$. Since $\partial\phi(0) \neq \emptyset$, we deduce from (2.2) that $0 \in \text{dom } \phi$ and that

$$(3.5) \quad (\forall \xi \in \mathbb{R}) \quad \sigma_{\Omega}(\xi) = \sup_{\nu \in \Omega} (\xi - 0)\nu \leq \phi(\xi) - \phi(0).$$

Consequently,

$$(3.6) \quad \text{dom } \phi \subset \text{dom } \sigma_{\Omega}.$$

Thus, in the case when $\Omega = \mathbb{R}$, Example 2.1(i) yields $\text{dom } \phi = \text{dom } \sigma_{\Omega} = \{0\}$ and we obtain $\phi = \phi(0) + \iota_{\{0\}} = \phi(0) + \sigma_{\Omega}$, hence (ii) with $\psi \equiv \phi(0)$. We henceforth assume that $\Omega \neq \mathbb{R}$ and set

$$(3.7) \quad (\forall \xi \in \mathbb{R}) \quad \varphi(\xi) = \begin{cases} \phi(\xi) - \phi(0) - \bar{\omega}\xi, & \text{if } \xi > 0 \text{ and } \bar{\omega} < +\infty; \\ \phi(\xi) - \phi(0) - \underline{\omega}\xi, & \text{if } \xi < 0 \text{ and } \underline{\omega} > -\infty; \\ 0, & \text{otherwise.} \end{cases}$$

Then Example 2.1(i) and (3.5) yield

$$(3.8) \quad \varphi \geq 0 = \varphi(0),$$

which also shows that φ is proper. In addition, we derive from Example 2.1(i) and (3.7) the following three possible expressions for φ .

(a) If $\underline{\omega} > -\infty$ and $\bar{\omega} < +\infty$, then σ_{Ω} is a finite continuous function and

$$(3.9) \quad (\forall \xi \in \mathbb{R}) \quad \varphi(\xi) = \phi(\xi) - \phi(0) - \sigma_{\Omega}(\xi).$$

(b) If $\underline{\omega} = -\infty$ and $\bar{\omega} < +\infty$, then

$$(3.10) \quad (\forall \xi \in \mathbb{R}) \quad \varphi(\xi) = \begin{cases} \phi(\xi) - \phi(0) - \bar{\omega}\xi, & \text{if } \xi > 0; \\ 0, & \text{otherwise.} \end{cases}$$

(c) If $\underline{\omega} > -\infty$ and $\bar{\omega} = +\infty$, then

$$(3.11) \quad (\forall \xi \in \mathbb{R}) \quad \varphi(\xi) = \begin{cases} \phi(\xi) - \phi(0) - \underline{\omega}\xi, & \text{if } \xi < 0; \\ 0, & \text{otherwise.} \end{cases}$$

Let us show that φ is lower semicontinuous. In case (a), this follows at once from the lower semicontinuity of ϕ and the continuity of σ_{Ω} . In cases (b) and (c), φ is clearly lower semicontinuous at every point $\xi \neq 0$ and, by (3.8), at 0 as well. Next, let us establish the convexity of φ . To this end, we set

$$(3.12) \quad (\forall \xi \in \mathbb{R}) \quad \bar{\varphi}(\xi) = \begin{cases} \phi(\xi) - \phi(0) - \bar{\omega}\xi, & \text{if } \xi > 0 \text{ and } \bar{\omega} < +\infty; \\ 0, & \text{otherwise,} \end{cases}$$

and

$$(3.13) \quad (\forall \xi \in \mathbb{R}) \quad \underline{\varphi}(\xi) = \begin{cases} \phi(\xi) - \phi(0) - \underline{\omega}\xi, & \text{if } \xi < 0 \text{ and } \underline{\omega} > -\infty; \\ 0, & \text{otherwise.} \end{cases}$$

By inspecting (3.7), (3.12), and (3.13) we learn that φ coincides with $\bar{\varphi}$ on $[0, +\infty[$ and with $\underline{\varphi}$ on $] -\infty, 0]$. Hence, (3.8) yields

$$(3.14) \quad \bar{\varphi} \geq 0 \quad \text{and} \quad \underline{\varphi} \geq 0,$$

and

$$(3.15) \quad \varphi = \max\{\underline{\varphi}, \bar{\varphi}\}.$$

Furthermore, since ϕ is convex, so are the functions $\xi \mapsto \phi(\xi) - \phi(0) - \bar{\omega}\xi$ and $\xi \mapsto \phi(\xi) - \phi(0) - \underline{\omega}\xi$, when $\bar{\omega} < +\infty$ and $\underline{\omega} > -\infty$, respectively. Therefore, it follows from (3.12), (3.13), and (3.14) that $\bar{\varphi}$ and $\underline{\varphi}$ are convex, and hence from (3.15) that φ is convex. We have thus shown that $\varphi \in \Gamma_0(\mathbb{R})$. We now claim that, for every $\xi \in \mathbb{R}$,

$$(3.16) \quad \phi(\xi) = \varphi(\xi) + \phi(0) + \sigma_\Omega(\xi).$$

We can establish this identity with the help of Example 2.1(i). In case (a), (3.16) follows at once from (3.9) since σ_Ω is finite. In case (b), (3.16) follows from (3.10) when $\xi \geq 0$, and from (3.5) when $\xi < 0$ since, in this case, $\sigma_\Omega(\xi) = +\infty$. Likewise, in case (c), (3.16) follows from (3.11) when $\xi \leq 0$, and from (3.5) when $\xi > 0$ since, in this case, $\sigma_\Omega(\xi) = +\infty$. Next, let us show that

$$(3.17) \quad 0 \in \text{int}(\text{dom } \phi - \text{dom } \sigma_\Omega).$$

In case (a), we have $\Omega = [\underline{\omega}, \bar{\omega}]$. Therefore $\text{dom } \sigma_\Omega = \mathbb{R}$ and (3.17) trivially holds. In case (b), we have $\Omega =]-\infty, \bar{\omega}]$ and, therefore, $\text{dom } \sigma_\Omega = [0, +\infty[$. This implies, via (3.6), that $\text{dom } \phi \subset [0, +\infty[$. Therefore, there exists $\nu \in \text{dom } \phi \cap]0, +\infty[$ since otherwise we would have $\text{dom } \phi = \{0\}$, which, in view of (2.2), would contradict the current working assumption that $\partial\phi(0) = \Omega \neq \mathbb{R}$. By convexity of ϕ , it follows that $[0, \nu] \subset \text{dom } \phi$ and, therefore, that $] -\infty, \nu] \subset \text{dom } \phi - \text{dom } \sigma_\Omega$. We thus obtain (3.17) in case (b); case (c) can be handled analogously. We can now appeal to [30, Theorem 23.8] to derive from (3.16), (3.17), and Example 2.1(ii) that

$$(3.18) \quad \Omega = \partial\phi(0) = \partial\varphi(0) + \partial\sigma_\Omega(0) = \partial\varphi(0) + \Omega$$

and, therefore, that $\partial\varphi(0) = \{0\}$ since $\Omega \neq \mathbb{R}$. In turn, upon invoking [30, Theorem 25.1], we conclude that φ is differentiable at 0 and that $\varphi'(0) = 0$. Altogether, we obtain (ii) by setting $\psi = \varphi + \phi(0)$. \square

Remark 3.4 A standard requirement for thresholders on \mathbb{R} is that they be nondecreasing functions [1, 31, 32, 37]. On the other hand, nonexpansivity is a key property to establish the convergence of iterative methods [11] and, in particular, in Proposition 1.1 [16] and Proposition 1.2 [14]. As seen in Proposition 2.4 and Definition 3.1, the nondecreasing and nonexpansive functions $\varrho: \mathbb{R} \rightarrow \mathbb{R}$ that vanish only on a closed interval $\Omega \subset \mathbb{R}$ coincide with the proximal thresholders on Ω . Hence, appealing to Theorem 3.3 and Lemma 2.3, we conclude that the operators that perform a componentwise nondecreasing and nonexpansive thresholding on $(\Omega_k)_{k \in \mathbb{K}}$ of those coefficients of the decomposition in $(e_k)_{k \in \mathbb{N}}$ indexed by \mathbb{K} are precisely the operators of the form prox_Ψ , where Ψ is as in (1.8).

Next, we provide a convenient decomposition rule for implementing proximal thresholders.

Proposition 3.5 *Let $\phi = \psi + \sigma_\Omega$, where $\psi \in \Gamma_0(\mathbb{R})$ and $\Omega \subset \mathbb{R}$ is a nonempty closed interval. Suppose that ψ is differentiable at 0 with $\psi'(0) = 0$. Then $\text{prox}_\phi = \text{prox}_\psi \circ \text{soft}_\Omega$.*

Proof. Fix ξ and π in \mathbb{R} . We have $0 \in \text{dom } \sigma_\Omega$ and, since ψ is differentiable at 0, $0 \in \text{int dom } \psi$. It therefore follows from (2.3) and [30, Theorem 23.8] that

$$\begin{aligned} \pi = \text{prox}_\phi \xi &\Leftrightarrow \xi - \pi \in \partial\phi(\pi) = \partial\psi(\pi) + \partial\sigma_\Omega(\pi) \\ (3.19) \quad &\Leftrightarrow (\exists \nu \in \partial\psi(\pi)) \quad \xi - (\pi + \nu) \in \partial\sigma_\Omega(\pi). \end{aligned}$$

Let us observe that, if $\nu \in \partial\psi(\pi)$, then, since $0 \in \text{Argmin } \psi$, (2.2) implies that $(0 - \pi)\nu + \psi(\pi) \leq \psi(0) \leq \psi(\pi) < +\infty$ and, in turn, that $\pi\nu \geq 0$. This shows that, if $\nu \in \partial\psi(\pi)$ and $\pi \neq 0$, then either $\pi > 0$ and $\nu \geq 0$, or $\pi < 0$ and $\nu \leq 0$; in turn, Lemma 2.1(ii) yields $\partial\sigma_\Omega(\pi) = \partial\sigma_\Omega(\pi + \nu)$. Consequently, if $\pi \neq 0$, we derive from (3.19) and Example 2.6(ii) that

$$\begin{aligned} \pi = \text{prox}_\phi \xi &\Rightarrow (\exists \nu \in \partial\psi(\pi)) \quad \xi - (\pi + \nu) \in \partial\sigma_\Omega(\pi + \nu) \\ &\Leftrightarrow (\exists \nu \in \partial\psi(\pi)) \quad \pi + \nu = \text{prox}_{\sigma_\Omega} \xi = \text{soft}_\Omega \xi \\ &\Leftrightarrow \text{soft}_\Omega \xi - \pi \in \partial\psi(\pi) \\ (3.20) \quad &\Leftrightarrow \pi = \text{prox}_\psi (\text{soft}_\Omega \xi). \end{aligned}$$

On the other hand, if $\pi = 0$, since $\partial\psi(0) = \{\psi'(0)\} = \{0\}$, we derive from (3.19), Example 2.1(ii), (1.10), and Lemma 2.2(i) that

$$\begin{aligned} \pi = \text{prox}_\phi \xi &\Rightarrow \xi \in \partial\sigma_\Omega(0) = \Omega \\ &\Rightarrow \text{soft}_\Omega \xi = 0 \\ (3.21) \quad &\Rightarrow \text{prox}_\psi (\text{soft}_\Omega \xi) = 0 = \pi. \end{aligned}$$

The proof is now complete. \square

In view of Proposition 3.5 and (1.10), the computation of the proximal thresholder $\text{prox}_{\psi+\sigma_\Omega}$ reduces to that of prox_ψ . By duality, we obtain a decomposition formula for those proximal operators that coincide with the identity on a closed interval Ω .

Proposition 3.6 *Let $\phi = \psi \square \iota_\Omega$, where $\psi \in \Gamma_0(\mathbb{R})$ and $\Omega \subset \mathbb{R}$ is a nonempty closed interval. Suppose that ψ^* is differentiable at 0 with $\psi^{*'}(0) = 0$. Then the following hold.*

- (i) $\text{prox}_\phi = P_\Omega + \text{prox}_\psi \circ \text{soft}_\Omega$.
- (ii) $(\forall \xi \in \mathbb{R}) \text{prox}_\phi \xi = \xi \Leftrightarrow \xi \in \Omega$.

Proof. It follows from [30, Theorem 16.4] that

$$(3.22) \quad \phi^* = \psi^* + \iota_\Omega^* = \psi^* + \sigma_\Omega.$$

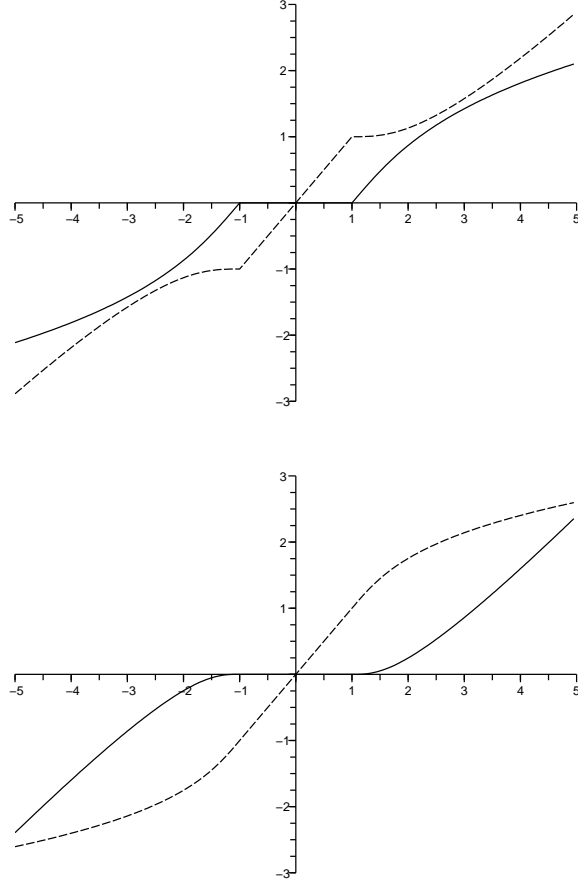


Figure 2: Graphs of the proximal threshold prox_ϕ (solid line) and its dual prox_{ϕ^*} (dashed line), where $\phi = \tau |\cdot|^p + |\cdot|$. Top: $\tau = 0.05$ and $p = 4$; Bottom: $\tau = 0.9$ and $p = 4/3$.

Note also that, since $\psi \in \Gamma_0(\mathbb{R})$, we have $\psi^* \in \Gamma_0(\mathbb{R})$ [30, Theorem 12.2]. (i): Fix $\xi \in \mathbb{R}$. Then, by Lemma 2.2(ii), (3.22), Proposition 3.5, and Example 2.6,

$$\begin{aligned}
 (3.23) \quad \text{prox}_\phi \xi &= \xi - \text{prox}_{\phi^*} \xi \\
 &= \xi - \text{prox}_{\psi^* + \sigma_\Omega} \xi \\
 &= \xi - \text{prox}_{\psi^*} (\text{prox}_{\sigma_\Omega} \xi) \\
 &= \xi - \text{prox}_{\sigma_\Omega} \xi + \text{prox}_\psi (\text{prox}_{\sigma_\Omega} \xi) \\
 &= \text{prox}_{\sigma_\Omega^*} \xi + \text{prox}_\psi (\text{prox}_{\sigma_\Omega} \xi) \\
 &= \text{prox}_{\iota_\Omega} \xi + \text{prox}_\psi (\text{prox}_{\sigma_\Omega} \xi) \\
 (3.24) \quad &= P_\Omega \xi + \text{prox}_\psi (\text{soft}_\Omega \xi).
 \end{aligned}$$

(ii): It follows from (3.22) and Theorem 3.3 that prox_{ϕ^*} is a proximal thresholder on Ω . Hence, we derive from (3.23) and (3.1) that $(\forall \xi \in \mathbb{R}) \text{prox}_{\phi} \xi = \xi \Leftrightarrow \text{prox}_{\phi^*} \xi = 0 \Leftrightarrow \xi \in \Omega$. \square

Examples of proximal thresholders (Proposition 3.5) and their duals (Proposition 3.6) are provided in Figs. 2 and 3 (see also Fig. 1) in the case when $\Omega = [-1, 1]$.

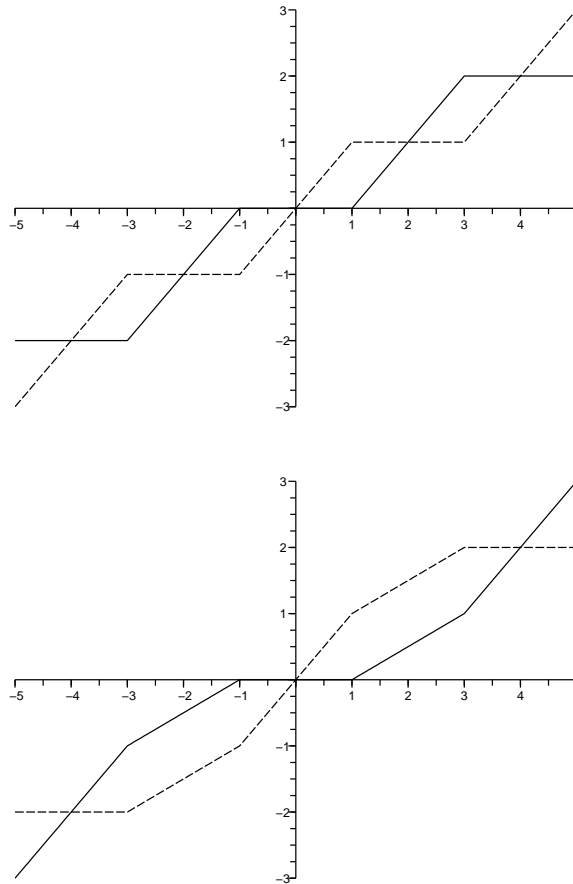


Figure 3: Graphs of the proximal thresholder prox_{ϕ} (solid line) and its dual prox_{ϕ^*} (dashed line), where $\phi = \psi + |\cdot|$. Top: $\psi = \iota_{[-2,2]}$; Bottom: $\psi: \xi \mapsto \xi^2/2$, if $|\xi| \leq 1$; $|\xi| - 1/2$, if $|\xi| > 1$, is the Huber function [25].

4 Iterative proximal thresholding

Let us start with some basic properties of Problem 1.3.

Proposition 4.1 *Problem 1.3 possesses at least one solution.*

Proof. Let Ψ be as in (1.8). We infer from the assumptions of Problem 1.3 and Lemma 2.3 that $\Psi \in \Gamma_0(\mathcal{H})$ and, in turn, that $\Phi + \Psi \in \Gamma_0(\mathcal{H})$. Hence, it suffices to show that $\Phi + \Psi$ is coercive [39, Theorem 2.5.1(ii)], i.e., since $\inf \Phi(\mathcal{H}) > -\infty$ by assumption (i) in Problem 1.3, that Ψ is coercive. For this purpose, let $\mathbf{x} = (\xi_k)_{k \in \mathbb{N}}$ denote a generic element in $\ell^2(\mathbb{N})$, and let

$$(4.1) \quad \Upsilon: \ell^2(\mathbb{N}) \rightarrow]-\infty, +\infty] : \mathbf{x} \mapsto \sum_{k \in \mathbb{N}} \psi_k(\xi_k) + \sum_{k \in \mathbb{K}} \sigma_{\Omega_k}(\xi_k).$$

Then, by Parseval's identity, it is enough to show that Υ is coercive. To this end, set $\mathbf{x}_{\mathbb{K}} = (\xi_k)_{k \in \mathbb{K}}$ and $\mathbf{x}_{\mathbb{L}} = (\xi_k)_{k \in \mathbb{L}}$, and denote by $\|\cdot\|_{\mathbb{K}}$ and $\|\cdot\|_{\mathbb{L}}$ the standard norms on $\ell^2(\mathbb{K})$ and $\ell^2(\mathbb{L})$, respectively. It follows from assumption (vi) in Problem 1.3 that there exists $\omega \in]0, +\infty[$ such that

$$(4.2) \quad [-\omega, \omega] \subset \bigcap_{k \in \mathbb{K}} \Omega_k.$$

Therefore, using (4.1), assumption (ii) in Problem 1.3, and Example 2.1(i), we obtain

$$(4.3) \quad \begin{aligned} (\forall \mathbf{x} \in \ell^2(\mathbb{N})) \quad \Upsilon(\mathbf{x}) &\geq \sum_{k \in \mathbb{K}} \sigma_{\Omega_k}(\xi_k) + \sum_{k \in \mathbb{L}} \psi_k(\xi_k) \\ &\geq \omega \sum_{k \in \mathbb{K}} |\xi_k| + \Upsilon_{\mathbb{L}}(\mathbf{x}_{\mathbb{L}}) \\ &\geq \omega \|\mathbf{x}_{\mathbb{K}}\|_{\mathbb{K}} + \Upsilon_{\mathbb{L}}(\mathbf{x}_{\mathbb{L}}). \end{aligned}$$

Now suppose that $\|\mathbf{x}\| = \sqrt{\|\mathbf{x}_{\mathbb{K}}\|_{\mathbb{K}}^2 + \|\mathbf{x}_{\mathbb{L}}\|_{\mathbb{L}}^2} \rightarrow +\infty$. Then (4.3) and assumption (v) in Problem 1.3 yield $\Upsilon(\mathbf{x}) \rightarrow +\infty$, as desired. \square

Proposition 4.2 *Let Ψ be as in (1.8), let $x \in \mathcal{H}$, and let $\gamma \in]0, +\infty[$. Then x is a solution to Problem 1.3 if and only if $x = \text{prox}_{\gamma\Psi}(x - \gamma\nabla\Phi(x))$.*

Proof. Since Problem 1.3 is equivalent to minimizing $\Phi + \Psi$, this is a standard characterization, see for instance [14, Proposition 3.1(iii)]. \square

Our algorithm for solving Problem 1.3 will be the following.

Algorithm 4.3 Fix $x_0 \in \mathcal{H}$ and set

$$(4.4) \quad \begin{aligned} (\forall n \in \mathbb{N}) \quad x_{n+1} &= x_n + \lambda_n \left(\sum_{k \in \mathbb{K}} \left(\alpha_{n,k} + \text{prox}_{\gamma_n \psi_k} \left(\text{soft}_{\gamma_n \Omega_k} \langle x_n - \gamma_n (\nabla\Phi(x_n) + b_n) \mid e_k \rangle \right) \right) e_k \right. \\ &\quad \left. + \sum_{k \in \mathbb{L}} \left(\alpha_{n,k} + \text{prox}_{\gamma_n \psi_k} \langle x_n - \gamma_n (\nabla\Phi(x_n) + b_n) \mid e_k \rangle \right) e_k - x_n \right), \end{aligned}$$

where:

- (i) $(\gamma_n)_{n \in \mathbb{N}}$ is a sequence in $]0, +\infty[$ such that $\inf_{n \in \mathbb{N}} \gamma_n > 0$ and $\sup_{n \in \mathbb{N}} \gamma_n < 2\beta$;
- (ii) $(\lambda_n)_{n \in \mathbb{N}}$ is a sequence in $]0, 1]$ such that $\inf_{n \in \mathbb{N}} \lambda_n > 0$;

- (iii) for every $n \in \mathbb{N}$, $(\alpha_{n,k})_{k \in \mathbb{N}}$ is a sequence in $\ell^2(\mathbb{N})$ such that $\sum_{n \in \mathbb{N}} \sqrt{\sum_{k \in \mathbb{N}} |\alpha_{n,k}|^2} < +\infty$;
- (iv) $(b_n)_{n \in \mathbb{N}}$ is a sequence in \mathcal{H} such that $\sum_{n \in \mathbb{N}} \|b_n\| < +\infty$.

Remark 4.4 Let us highlight some features of Algorithm 4.3.

- The set \mathbb{K} contains the indices of those coefficients of the decomposition in $(e_k)_{k \in \mathbb{N}}$ that are thresholded.
- The terms $\alpha_{n,k}$ and b_n stand for some numerical tolerance in the implementation of $\text{prox}_{\gamma_n \psi_k}$ and the computation of $\nabla \Phi(x_n)$, respectively.
- The parameters λ_n and γ_n provide added flexibility to the algorithm and can be used to improve its convergence profile.
- The operator $\text{soft}_{\gamma_n \Omega_k}$ is given explicitly in (1.10).

Our main convergence result can now be stated.

Theorem 4.5 *Every sequence generated by Algorithm 4.3 converges strongly to a solution to Problem 1.3.*

Proof. Hereafter, the arrow \rightharpoonup stands for weak convergence, $(x_n)_{n \in \mathbb{N}}$ is a sequence generated by Algorithm 4.3, and we define

$$(4.5) \quad (\forall k \in \mathbb{N}) \quad \phi_k = \begin{cases} \psi_k + \sigma_{\Omega_k}, & \text{if } k \in \mathbb{K}; \\ \psi_k, & \text{if } k \in \mathbb{L}. \end{cases}$$

It follows from the assumptions on $(\psi_k)_{k \in \mathbb{N}}$ in Problem 1.3 that $(\forall k \in \mathbb{N}) \psi'_k(0) = 0$. Therefore, for every $n \in \mathbb{N}$, Theorem 3.3 implies that

$$(4.6) \quad \text{for every } k \text{ in } \mathbb{K}, \text{prox}_{\gamma_n \phi_k} \text{ is a proximal thresholder on } \gamma_n \Omega_k,$$

while Proposition 3.5 supplies

$$(4.7) \quad (\forall k \in \mathbb{K}) \quad \text{prox}_{\gamma_n \phi_k} = \text{prox}_{\gamma_n \psi_k + \gamma_n \sigma_{\Omega_k}} = \text{prox}_{\gamma_n \psi_k + \sigma_{(\gamma_n \Omega_k)}} = \text{prox}_{\gamma_n \psi_k} \circ \text{soft}_{\gamma_n \Omega_k}.$$

Thus, (4.4) can be rewritten as

$$(4.8) \quad x_{n+1} = x_n + \lambda_n \left(\sum_{k \in \mathbb{N}} (\alpha_{n,k} + \text{prox}_{\gamma_n \phi_k} \langle x_n - \gamma_n (\nabla \Phi(x_n) + b_n) \mid e_k \rangle) e_k - x_n \right).$$

Now, let Ψ be as in (1.8), i.e., $\Psi = \sum_{k \in \mathbb{N}} \phi_k(\langle \cdot \mid e_k \rangle)$, and set $(\forall n \in \mathbb{N}) a_n = \sum_{k \in \mathbb{N}} \alpha_{n,k} e_k$. Then it follows from (4.5) and Lemma 2.3 that $\Psi \in \Gamma_0(\mathcal{H})$ and that (4.8) can be rewritten as

$$(4.9) \quad x_{n+1} = x_n + \lambda_n \left(\text{prox}_{\gamma_n \Psi} (x_n - \gamma_n (\nabla \Phi(x_n) + b_n)) + a_n - x_n \right).$$

Consequently, since Proposition 4.1 asserts that $\Phi + \Psi$ possesses a minimizer, we derive from assumptions (i)–(iv) in Algorithm 4.3 and [14, Theorem 3.4] that

$$(4.10) \quad (x_n)_{n \in \mathbb{N}} \text{ converges weakly to a solution } x \text{ to Problem 1.3}$$

and that

$$(4.11) \quad \sum_{n \in \mathbb{N}} \|x_n - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x_n))\|^2 < +\infty \quad \text{and} \quad \sum_{n \in \mathbb{N}} \|\nabla \Phi(x_n) - \nabla \Phi(x)\|^2 < +\infty.$$

Hence, it follows from Lemma 2.2(iii) and assumption (i) in Algorithm 4.3 that

$$(4.12) \quad \begin{aligned} & \frac{1}{2} \sum_{n \in \mathbb{N}} \|x_n - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x))\|^2 \\ & \leq \sum_{n \in \mathbb{N}} \|x_n - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x_n))\|^2 \\ & \quad + \sum_{n \in \mathbb{N}} \|\text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x_n)) - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x))\|^2 \\ & \leq \sum_{n \in \mathbb{N}} \|x_n - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x_n))\|^2 + \sum_{n \in \mathbb{N}} \gamma_n^2 \|\nabla \Phi(x_n) - \nabla \Phi(x)\|^2 \\ & \leq \sum_{n \in \mathbb{N}} \|x_n - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x_n))\|^2 + 4\beta^2 \sum_{n \in \mathbb{N}} \|\nabla \Phi(x_n) - \nabla \Phi(x)\|^2 \\ & < +\infty. \end{aligned}$$

Now define

$$(4.13) \quad (\forall n \in \mathbb{N}) \quad v_n = x_n - x \quad \text{and} \quad h_n = x - \gamma_n \nabla \Phi(x).$$

On the one hand, we derive from (4.10) that

$$(4.14) \quad v_n \rightharpoonup 0$$

and, on the other hand, from (4.12) and Proposition 4.2 that

$$(4.15) \quad \sum_{n \in \mathbb{N}} \|v_n - \text{prox}_{\gamma_n \Psi}(v_n + h_n) + \text{prox}_{\gamma_n \Psi} h_n\|^2 = \sum_{n \in \mathbb{N}} \|x_n - \text{prox}_{\gamma_n \Psi}(x_n - \gamma_n \nabla \Phi(x))\|^2 < +\infty.$$

By Parseval's identity, to establish that $\|v_n\| = \|x_n - x\| \rightarrow 0$, we must show that

$$(4.16) \quad \sum_{k \in \mathbb{K}} |\nu_{n,k}|^2 \rightarrow 0 \quad \text{and} \quad \sum_{k \in \mathbb{L}} |\nu_{n,k}|^2 \rightarrow 0, \quad \text{where} \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{N}) \quad \nu_{n,k} = \langle v_n | e_k \rangle.$$

To this end, it is useful to set, for every $n \in \mathbb{N}$ and $k \in \mathbb{N}$, $\eta_{n,k} = \langle h_n | e_k \rangle$ and observe that (4.15), Parseval's identity, and Lemma 2.3 imply that

$$(4.17) \quad \sum_{k \in \mathbb{N}} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|^2 \rightarrow 0.$$

In addition, if we set $r = 2\beta\nabla\Phi(x)$ and, for every $k \in \mathbb{N}$, $\xi_k = \langle x \mid e_k \rangle$ and $\rho_k = \langle r \mid e_k \rangle$, then we derive from (4.13) and assumption (i) in Algorithm 4.3 that

$$(4.18) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{N}) \quad |\eta_{n,k}|^2/2 \leq |\xi_k|^2 + \gamma_n^2 |\langle \nabla\Phi(x) \mid e_k \rangle|^2 \leq |\xi_k|^2 + |\rho_k|^2.$$

To establish (4.16), let us first show that $\sum_{k \in \mathbb{K}} |\nu_{n,k}|^2 \rightarrow 0$. Assumption (vi) in Problem 1.3 asserts that there exists $\omega \in]0, +\infty[$ such that

$$(4.19) \quad [-\omega, \omega] \subset \bigcap_{k \in \mathbb{K}} \Omega_k.$$

Now set $\delta = \gamma\omega$, where $\gamma = \inf_{n \in \mathbb{N}} \gamma_n$. Then it follows from assumption (i) in Algorithm 4.3 that $\delta > 0$ and from (4.19) that

$$(4.20) \quad [-\delta, \delta] \subset \bigcap_{n \in \mathbb{N}} \bigcap_{k \in \mathbb{K}} \gamma_n \Omega_k.$$

On the other hand, (4.18) yields

$$(4.21) \quad \sum_{k \in \mathbb{K}} \sup_{n \in \mathbb{N}} |\eta_{n,k}|^2/2 \leq \sum_{k \in \mathbb{K}} (|\xi_k|^2 + |\rho_k|^2) = \|x\|^2 + \|r\|^2 < +\infty.$$

Hence, there exists a finite set $\mathbb{K}_1 \subset \mathbb{K}$ such that

$$(4.22) \quad (\forall n \in \mathbb{N}) \quad \sum_{k \in \mathbb{K}_2} |\eta_{n,k}|^2 \leq \delta^2/4, \quad \text{where } \mathbb{K}_2 = \mathbb{K} \setminus \mathbb{K}_1.$$

In view of (4.14), we have $\sum_{k \in \mathbb{K}_1} |\nu_{n,k}|^2 \rightarrow 0$. Let us now show that $\sum_{k \in \mathbb{K}_2} |\nu_{n,k}|^2 \rightarrow 0$. Note that (4.20) and (4.22) yield

$$(4.23) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{K}_2) \quad \eta_{n,k} \in [-\delta/2, \delta/2] \subset \gamma_n \Omega_k.$$

Therefore, (4.6) implies that

$$(4.24) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{K}_2) \quad \text{prox}_{\gamma_n \phi_k} \eta_{n,k} = 0.$$

Let us define

$$(4.25) \quad (\forall n \in \mathbb{N}) \quad \mathbb{K}_{21,n} = \{k \in \mathbb{K}_2 \mid \nu_{n,k} + \eta_{n,k} \in \gamma_n \Omega_k\}.$$

Then, invoking (4.6) once again, we obtain

$$(4.26) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{K}_{21,n}) \quad \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) = 0$$

which, combined with (4.24), yields

$$(4.27) \quad \begin{aligned} (\forall n \in \mathbb{N}) \quad \sum_{k \in \mathbb{K}_{21,n}} |\nu_{n,k}|^2 &= \sum_{k \in \mathbb{K}_{21,n}} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|^2 \\ &\leq \sum_{k \in \mathbb{N}} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|^2. \end{aligned}$$

Consequently, it results from (4.17) that $\sum_{k \in \mathbb{K}_{21,n}} |\nu_{n,k}|^2 \rightarrow 0$. Next, let us set

$$(4.28) \quad (\forall n \in \mathbb{N}) \quad \mathbb{K}_{22,n} = \mathbb{K}_2 \setminus \mathbb{K}_{21,n}$$

and show that $\sum_{k \in \mathbb{K}_{22,n}} |\nu_{n,k}|^2 \rightarrow 0$. It follows from (4.28), (4.25), and (4.20) that

$$(4.29) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{K}_{22,n}) \quad \nu_{n,k} + \eta_{n,k} \notin \gamma_n \Omega_k \supset [-\delta, \delta].$$

Hence, appealing to (4.23), we obtain

$$(4.30) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{K}_{22,n}) \quad |\nu_{n,k} + \eta_{n,k}| \geq \delta \geq |\eta_{n,k}| + \delta/2.$$

Now take $n \in \mathbb{N}$ and $k \in \mathbb{K}_{22,n}$. We derive from (4.24) and Lemma 2.2(ii) that

$$(4.31) \quad \begin{aligned} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}| \\ = |(\nu_{n,k} + \eta_{n,k}) - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) - \eta_{n,k}| \\ = |\text{prox}_{(\gamma_n \phi_k)^*}(\nu_{n,k} + \eta_{n,k}) - \eta_{n,k}|. \end{aligned}$$

However, it results from (4.20), (4.6), and Proposition 3.2 that $\text{prox}_{(\gamma_n \phi_k)^*}(\pm\delta) = \pm\delta$. We consider two cases. First, if $\nu_{n,k} + \eta_{n,k} \geq 0$ then, since $\text{prox}_{(\gamma_n \phi_k)^*}$ is nondecreasing by Proposition 2.4, (4.30) yields $\nu_{n,k} + \eta_{n,k} \geq \delta$ and

$$(4.32) \quad \text{prox}_{(\gamma_n \phi_k)^*}(\nu_{n,k} + \eta_{n,k}) \geq \text{prox}_{(\gamma_n \phi_k)^*} \delta = \delta \geq \eta_{n,k} + \delta/2.$$

Likewise, if $\nu_{n,k} + \eta_{n,k} \leq 0$, then (4.30) yields $\nu_{n,k} + \eta_{n,k} \leq -\delta$ and

$$(4.33) \quad \text{prox}_{(\gamma_n \phi_k)^*}(\nu_{n,k} + \eta_{n,k}) \leq \text{prox}_{(\gamma_n \phi_k)^*}(-\delta) = -\delta \leq \eta_{n,k} - \delta/2.$$

Altogether, we derive from (4.32) and (4.33) that

$$(4.34) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{K}_{22,n}) \quad |\text{prox}_{(\gamma_n \phi_k)^*}(\nu_{n,k} + \eta_{n,k}) - \eta_{n,k}| \geq \delta/2.$$

In turn, (4.31) yields

$$(4.35) \quad (\forall n \in \mathbb{N}) \quad \sum_{k \in \mathbb{K}_{22,n}} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|^2 \geq \text{card}(\mathbb{K}_{22,n}) \delta^2/4.$$

However, it follows from (4.17) that, for n sufficiently large,

$$(4.36) \quad \sum_{k \in \mathbb{N}} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|^2 \leq \delta^2/5.$$

Thus, for n sufficiently large, $\mathbb{K}_{22,n} = \emptyset$. We conclude from this first part of the proof that $\sum_{k \in \mathbb{K}} |\nu_{n,k}|^2 \rightarrow 0$.

In order to obtain (4.16), we must now show that $\sum_{k \in \mathbb{L}} |\nu_{n,k}|^2 \rightarrow 0$. We infer from (4.14) that $(v_n)_{n \in \mathbb{N}}$ is bounded, hence

$$(4.37) \quad \sup_{n \in \mathbb{N}} \sum_{k \in \mathbb{L}} |\nu_{n,k}|^2 \leq \sup_{n \in \mathbb{N}} \|v_n\|^2 \leq \rho^2/4,$$

for some $\rho \in]0, +\infty[$. Now define

$$(4.38) \quad \mathbb{L}_1 = \{k \in \mathbb{L} \mid (\exists n \in \mathbb{N}) |\eta_{n,k}| \geq \rho/2\}.$$

Then we derive from (4.18) that

$$(4.39) \quad (\forall k \in \mathbb{L}_1)(\exists n \in \mathbb{N}) \quad |\xi_k|^2 + |\rho_k|^2 \geq |\eta_{n,k}|^2/2 \geq \rho^2/8.$$

Consequently, we have

$$(4.40) \quad +\infty > \|x\|^2 + \|r\|^2 \geq \sum_{k \in \mathbb{L}_1} (|\xi_k|^2 + |\rho_k|^2) \geq (\text{card } \mathbb{L}_1)\rho^2/8$$

and therefore $\text{card}(\mathbb{L}_1) < +\infty$. In turn, it results from (4.14) that $\sum_{k \in \mathbb{L}_1} |\nu_{n,k}|^2 \rightarrow 0$. Hence, to obtain $\sum_{k \in \mathbb{L}} |\nu_{n,k}|^2 \rightarrow 0$, it remains to show that $\sum_{k \in \mathbb{L}_2} |\nu_{n,k}|^2 \rightarrow 0$, where $\mathbb{L}_2 = \mathbb{L} \setminus \mathbb{L}_1$. In view of (4.38) and (4.37), we have

$$(4.41) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{L}_2) \quad |\eta_{n,k}| < \rho/2 \quad \text{and} \quad |\nu_{n,k} + \eta_{n,k}| \leq |\nu_{n,k}| + |\eta_{n,k}| < \rho.$$

On the other hand, assumption (iv) in Problem 1.3 asserts that there exists $\theta \in]0, +\infty[$ such that

$$(4.42) \quad \inf_{n \in \mathbb{N}} \inf_{k \in \mathbb{L}_2} \inf_{0 < |\xi| \leq \rho} (\gamma_n \psi_k)''(\xi) \geq \gamma \inf_{k \in \mathbb{L}_2} \inf_{0 < |\xi| \leq \rho} \psi_k''(\xi) \geq \gamma\theta.$$

It therefore follows from assumptions (ii) and (iii) in Problem 1.3, Proposition 2.8, and (4.5) that

$$(4.43) \quad \begin{aligned} (\forall n \in \mathbb{N})(\forall k \in \mathbb{L}_2) \quad |\nu_{n,k}| &\leq |\nu_{n,k} - \text{prox}_{\gamma_n \psi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \psi_k} \eta_{n,k}| \\ &\quad + |\text{prox}_{\gamma_n \psi_k}(\nu_{n,k} + \eta_{n,k}) - \text{prox}_{\gamma_n \psi_k} \eta_{n,k}| \\ &\leq |\nu_{n,k} - \text{prox}_{\gamma_n \psi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \psi_k} \eta_{n,k}| \\ &\quad + |\nu_{n,k}|/(1 + \gamma\theta) \\ &= |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}| \\ &\quad + |\nu_{n,k}|/(1 + \gamma\theta). \end{aligned}$$

Consequently, upon setting $\mu = 1 + 1/(\gamma\theta)$, we obtain

$$(4.44) \quad (\forall n \in \mathbb{N})(\forall k \in \mathbb{L}_2) \quad |\nu_{n,k}| \leq \mu |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|.$$

In turn,

$$(4.45) \quad (\forall n \in \mathbb{N}) \quad \sum_{k \in \mathbb{L}_2} |\nu_{n,k}|^2 \leq \mu^2 \sum_{k \in \mathbb{L}_2} |\nu_{n,k} - \text{prox}_{\gamma_n \phi_k}(\nu_{n,k} + \eta_{n,k}) + \text{prox}_{\gamma_n \phi_k} \eta_{n,k}|^2.$$

Hence, (4.17) forces $\sum_{k \in \mathbb{L}_2} |\nu_{n,k}|^2 \rightarrow 0$, as desired. \square

Remark 4.6 An important aspect of Theorem 4.5 is that it provides a *strong* convergence result. Indeed, in general, only weak convergence can be claimed for forward-backward methods [14, 36] (see [3], [4], [14, Remark 5.12], and [23] for explicit constructions in which strong convergence fails). In addition, the standard sufficient conditions for strong convergence in this type of algorithm (see [11, Remark 6.6] and [14, Theorem 3.4(iv)]) are not satisfied in Problem 1.3. Further aspects of the relevance of strong convergence in proximal methods are discussed in [23, 24].

Remark 4.7 Let T be a nonzero bounded linear operator from \mathcal{H} to a real Hilbert space \mathcal{G} , let $z \in \mathcal{G}$, and let τ and ω be in $]0, +\infty[$. Specializing Theorem 4.5 to the case when $\Phi: x \mapsto \|Tx - z\|^2/2$ and either

$$(4.46) \quad \mathbb{K} = \emptyset \quad \text{and} \quad (\forall k \in \mathbb{L}) \quad \psi_k = \tau_k |\cdot|^p, \quad \text{where} \quad p \in]1, 2] \quad \text{and} \quad \tau_k \in [\tau, +\infty[,$$

or

$$(4.47) \quad \mathbb{L} = \emptyset \quad \text{and} \quad (\forall k \in \mathbb{K}) \quad \psi_k = 0 \quad \text{and} \quad \Omega_k = [-\omega_k, \omega_k], \quad \text{where} \quad \omega_k \in [\omega, +\infty[,$$

yields [14, Corollary 5.19]. If we further impose $\lambda_n \equiv 1$, $\|T\| < 1$, $\gamma_n \equiv 1$, $\alpha_{n,k} \equiv 0$, and $b_n \equiv 0$, we obtain [16, Theorem 3.1].

5 Applications to sparse signal recovery

5.1 A special case of Problem 1.3

In (1.3), a single observation of the original signal \bar{x} is available. In certain problems, q such noisy linear observations are available, say $z_i = T_i \bar{x} + v_i$ ($1 \leq i \leq q$), which leads to the weighted least-squares data fidelity term $x \mapsto \sum_{i=1}^q \mu_i \|T_i x - z_i\|^2$; see [10] and the references therein. Furthermore, signal recovery problems are typically accompanied with convex constraints that confine \bar{x} to some closed convex subsets $(S_i)_{1 \leq i \leq m}$ of \mathcal{H} . These constraints can be aggregated via the cost function $x \mapsto \sum_{i=1}^m \nu_i d_{S_i}^2(x)$; see [9, 26] and the references therein. On the other hand, a common approach to penalize the coefficients of an orthonormal basis decomposition is to use power functions, e.g., [1, 7, 16]. Moreover, we aim at promoting sparsity of a solution $x \in \mathcal{H}$ with respect to $(e_k)_{k \in \mathbb{N}}$ in the sense that, for every k in \mathbb{K} , we wish to set to 0 the coefficient $\langle x | e_k \rangle$ if it lies in the interval Ω_k . Altogether, these considerations suggest the following formulation.

Problem 5.1 For every $i \in \{1, \dots, q\}$, let $\mu_i \in]0, +\infty[$, let T_i be a nonzero bounded linear operator from \mathcal{H} to a real Hilbert space \mathcal{G}_i , and let $z_i \in \mathcal{G}_i$. For every $i \in \{1, \dots, m\}$, let $\nu_i \in]0, +\infty[$ and let S_i be a nonempty closed and convex subset of \mathcal{H} . Furthermore, let $(p_{k,l})_{0 \leq l \leq L_k}$ be distinct real numbers in $]1, +\infty[$, let $(\tau_{k,l})_{0 \leq l \leq L_k}$ be real numbers in $[0, +\infty[$, and let $l_k \in \{0, \dots, L_k\}$ satisfy $p_{k,l_k} = \min_{0 \leq l \leq L_k} p_{k,l}$, where $(L_k)_{k \in \mathbb{N}}$ is a sequence in \mathbb{N} . Finally, let $\mathbb{K} \subset \mathbb{N}$, let $\mathbb{L} = \mathbb{N} \setminus \mathbb{K}$, and let $(\Omega_k)_{k \in \mathbb{K}}$ be a sequence of closed intervals in \mathbb{R} . The objective is to

$$(5.1) \quad \underset{x \in \mathcal{H}}{\text{minimize}} \quad \frac{1}{2} \sum_{i=1}^q \mu_i \|T_i x - z_i\|^2 + \frac{1}{2} \sum_{i=1}^m \nu_i d_{S_i}^2(x) + \sum_{k \in \mathbb{N}} \sum_{l=0}^{L_k} \tau_{k,l} |\langle x | e_k \rangle|^{p_{k,l}} + \sum_{k \in \mathbb{K}} \sigma_{\Omega_k}(\langle x | e_k \rangle),$$

under the following assumptions:

- (i) $\inf_{k \in \mathbb{L}} \tau_{k,l_k} > 0$;
- (ii) $\inf_{k \in \mathbb{L}} p_{k,l_k} > 1$;
- (iii) $\sup_{k \in \mathbb{L}} p_{k,l_k} \leq 2$;

(iv) $0 \in \text{int} \bigcap_{k \in \mathbb{K}} \Omega_k$.

Proposition 5.2 *Problem 5.1 is a special case of Problem 1.3.*

Proof. First, we observe that (5.1) corresponds to (1.6) where

$$(5.2) \quad \Phi: x \mapsto \frac{1}{2} \sum_{i=1}^q \mu_i \|T_i x - z_i\|^2 + \frac{1}{2} \sum_{i=1}^m \nu_i d_{S_i}^2(x) \quad \text{and} \quad (\forall k \in \mathbb{N}) \quad \psi_k: \xi \mapsto \sum_{l=0}^{L_k} \tau_{k,l} |\xi|^{p_{k,l}}.$$

Hence, Φ is a finite continuous convex function with Fréchet gradient

$$(5.3) \quad \nabla \Phi: x \mapsto \sum_{i=1}^q \mu_i T_i^*(T_i x - z_i) + \sum_{i=1}^m \nu_i (x - P_i x),$$

where P_i is the projection operator onto S_i . Therefore, since the operators $(\text{Id} - P_i)_{1 \leq i \leq m}$ are nonexpansive, it follows that assumption (i) in Problem 1.3 is satisfied with $1/\beta = \sum_{i=1}^q \mu_i \|T_i\|^2 + \sum_{i=1}^m \nu_i$. Moreover, the functions $(\psi_k)_{k \in \mathbb{N}}$ are in $\Gamma_0(\mathbb{R})$ and satisfy assumptions (ii) and (iii) in Problem 1.3.

Let us now turn to assumption (iv) in Problem 1.3. Fix $\rho \in]0, +\infty[$ and set $\tau = \inf_{k \in \mathbb{L}} \tau_{k,l_k}$, $p = \inf_{k \in \mathbb{L}} p_{k,l_k}$, and $\theta = \tau p(p-1) \min\{1, 1/\rho\}$. Then it follows from (i), (ii), and (iii) that $\theta > 0$ and that

$$(5.4) \quad \begin{aligned} \inf_{k \in \mathbb{L}} \inf_{0 < |\xi| \leq \rho} \psi_k''(\xi) &= \inf_{k \in \mathbb{L}} \inf_{0 < |\xi| \leq \rho} \sum_{l=0}^{L_k} \tau_{k,l} p_{k,l} (p_{k,l} - 1) |\xi|^{p_{k,l}-2} \\ &\geq \inf_{k \in \mathbb{L}} \tau_{k,l_k} p_{k,l_k} (p_{k,l_k} - 1) \inf_{0 < \xi \leq \rho} \xi^{p_{k,l_k}-2} \\ &\geq \tau p(p-1) \inf_{k \in \mathbb{L}} \inf_{0 < \xi \leq \rho} \xi^{p_{k,l_k}-2} \\ &\geq \theta, \end{aligned}$$

which shows that (1.7) is satisfied.

It remains to check assumption (v) in Problem 1.3. To this end, let $\|\cdot\|_{\mathbb{L}}$ denote the standard norm on $\ell^2(\mathbb{L})$, take $\mathbf{x} = (\xi_k)_{k \in \mathbb{L}} \in \ell^2(\mathbb{L})$ such that $\|\mathbf{x}\|_{\mathbb{L}} \geq 1$, and set $(\eta_k)_{k \in \mathbb{L}} = \mathbf{x}/\|\mathbf{x}\|_{\mathbb{L}}$. Then, for every $k \in \mathbb{L}$, $|\eta_k| \leq 1$ and, since $p_{k,l_k} \in]1, 2]$, we have $|\eta_k|^{p_{k,l_k}} \geq |\eta_k|^2$. Consequently,

$$(5.5) \quad \begin{aligned} \Upsilon_{\mathbb{L}}(\mathbf{x}) &= \sum_{k \in \mathbb{L}} \sum_{l=0}^{L_k} \tau_{k,l} |\xi_k|^{p_{k,l}} \geq \sum_{k \in \mathbb{L}} \tau_{k,l_k} |\xi_k|^{p_{k,l_k}} \\ &\geq \tau \sum_{k \in \mathbb{L}} |\xi_k|^{p_{k,l_k}} = \tau \sum_{k \in \mathbb{L}} \|\mathbf{x}\|_{\mathbb{L}}^{p_{k,l_k}} |\eta_k|^{p_{k,l_k}} \\ &\geq \tau \sum_{k \in \mathbb{L}} \|\mathbf{x}\|_{\mathbb{L}}^{p_{k,l_k}} |\eta_k|^2 = \tau \sum_{k \in \mathbb{L}} \|\mathbf{x}\|_{\mathbb{L}}^{p_{k,l_k}-2} |\xi_k|^2 \\ &\geq \tau \|\mathbf{x}\|_{\mathbb{L}}^{-1} \sum_{k \in \mathbb{L}} |\xi_k|^2 = \tau \|\mathbf{x}\|_{\mathbb{L}}. \end{aligned}$$

We conclude that $\Upsilon_{\mathbb{L}}(\mathbf{x}) \rightarrow +\infty$ as $\|\mathbf{x}\|_{\mathbb{L}} \rightarrow +\infty$. \square

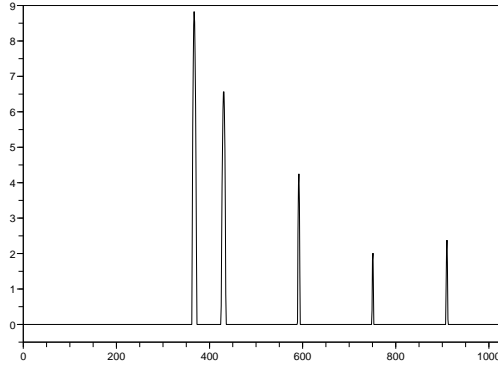


Figure 4: Original signal – first example.

5.2 First example

Our first example concerns the simulated X-ray fluorescence spectrum \bar{x} displayed in Fig. 4, which is often used to test restoration methods. The underlying Hilbert space is $\mathcal{H} = \ell^2(\mathbb{N})$. The measured signal z shown in Fig. 5 has undergone blurring by the limited resolution of the spectrometer and further corruption by addition of noise. In the underlying Hilbert space $\mathcal{H} = \ell^2(\mathbb{N})$, this process is modeled by $z = T\bar{x} + v$, where $T: \mathcal{H} \rightarrow \mathcal{H}$ is the operator of convolution with a truncated Gaussian kernel. The noise samples are uncorrelated and drawn from a Gaussian population with mean zero and standard deviation 0.15. The original signal \bar{x} has support $\{0, \dots, N - 1\}$ ($N = 1024$), takes on nonnegative values, and possesses a sparse structure. These features can be promoted in Problem 5.1 by letting $(e_k)_{k \in \mathbb{N}}$ be the canonical orthonormal basis of \mathcal{H} , and setting $\mathbb{K} = \mathbb{N}$, $\tau_{k,l} \equiv 0$, and

$$(5.6) \quad (\forall k \in \mathbb{N}) \quad \Omega_k = \begin{cases}]-\infty, \omega], & \text{if } 0 \leq k \leq N - 1; \\ \mathbb{R}, & \text{otherwise,} \end{cases}$$

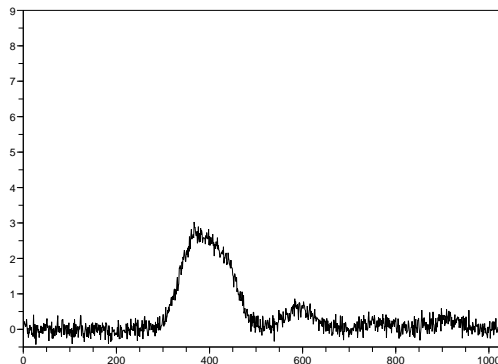


Figure 5: Degraded signal – first example.

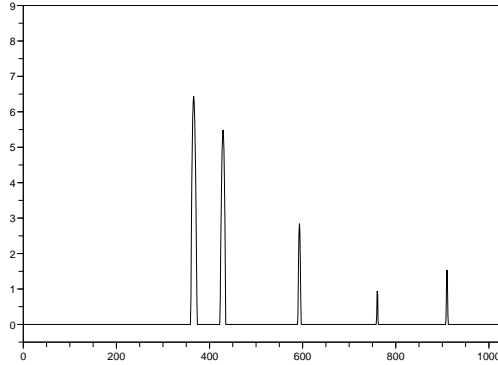


Figure 6: Signal restored by the proposed algorithm – first example.

where the one-sided thresholding level is set to $\omega = 0.01$. On the other hand, using the methodology described in [35], the above information about the noise can be used to construct the constraint sets $S_1 = \{x \in \mathcal{H} \mid \|Tx - z\| \leq \delta_1\}$ and $S_2 = \bigcap_{l=1}^{N-1} \{x \in \mathcal{H} \mid |\widehat{Tx}(l/N) - \widehat{z}(l/N)| \leq \delta_2\}$, where $\widehat{a}: \nu \mapsto \sum_{k=0}^{+\infty} \langle a \mid e_k \rangle \exp(-i2\pi k\nu)$ designates the Fourier transform of $a \in \mathcal{H}$. The bounds δ_1 and δ_2 have been determined so as to guarantee that \bar{x} lies in S_1 and in S_2 with a 99 percent confidence level (see [13] for details). Finally, we set $q = 0$ and $\nu_1 = \nu_2 = 1$ in (5.1) (the computation of the projectors P_1 and P_2 required in (5.3) is detailed in [35]). The solution produced by Algorithm 4.3 is shown in Fig. 6. It is of much better quality than the restorations obtained in [12] and [35] via alternative methods.

5.3 Second example

We provide a wavelet deconvolution example in $\mathcal{H} = \mathbb{L}^2(\mathbb{R})$. The original signal \bar{x} is the classical “bumps” signal [38] displayed in Fig. 7. The degraded version shown in Fig. 8 is $z_1 = T_1\bar{x} + v_1$, where T_1 models convolution with a uniform kernel and v_1 is a realization of a zero-mean white

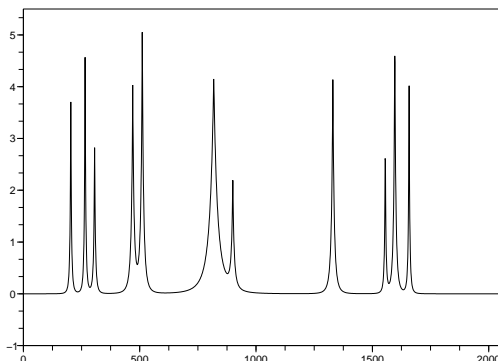


Figure 7: Original signal – second example.

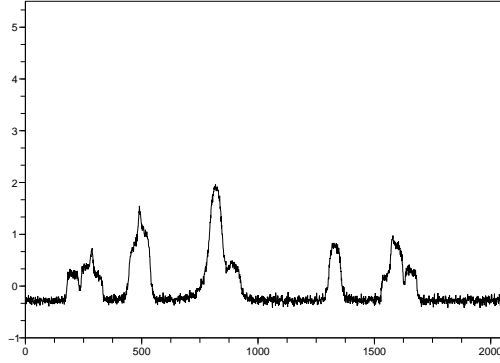


Figure 8: Degraded signal – second example.

Gaussian noise. The basis $(e_k)_{k \in \mathbb{N}}$ is an orthonormal wavelet symlet basis with 8 vanishing moments [15]. Such wavelet bases are known to provide sparse representations for a wide class of signals [20] such as this standard test signal. Note that there exists a strong connection between Problem 5.1 and maximum *a posteriori* techniques for estimating \bar{x} in the presence of white Gaussian noise. In particular, setting $q = 1$, $m = 0$, $\mathbb{K} = \emptyset$ and $L_k \equiv 0$, and using suitably subband-adapted values of $p_{k,0}$ and $\tau_{k,0}$ amounts to fitting an appropriate generalized Gaussian prior distribution to the wavelet coefficients in each subband [1]. Such a statistical modeling is commonly used in wavelet-based estimation, where values of $p_{k,0}$ close to 2 may provide a good model at coarse resolution levels, whereas values close to 1 should preferably be used at finer resolutions. The setting of the more general model we adopt here is the following: in Problem 5.1, \mathbb{K} and \mathbb{L} are the index sets of the detail and approximation coefficients [27], respectively, and

- $(\forall k \in \mathbb{K}) \Omega_k = [-0.0023, 0.0023]$, $L_k = 1$, $(p_{k,0}, p_{k,1}) = (2, 4)$, $(\tau_{k,0}, \tau_{k,1}) = (0.0052, 0.0001)$.
- $(\forall k \in \mathbb{L}) L_k = 0$, $p_{k,0} = 2$, $\tau_{k,0} = 0.00083$.

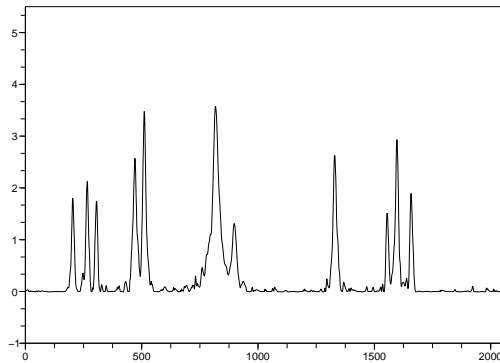


Figure 9: Signal restored by the proposed algorithm – second example.

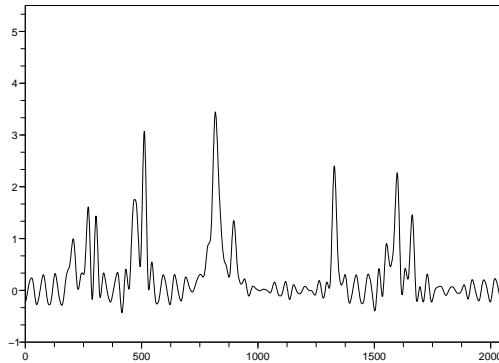


Figure 10: Signal restored by solving (1.3) – second example.

In addition, we set $q = 1$, $\mu_1 = 1$, $m = 1$, $\nu_1 = 1$, and $S_1 = \{x \in \mathcal{H} \mid x \geq 0\}$ (nonnegativity constraint). The solution x produced by Algorithm 4.3 is shown in Fig. 9. The estimation error is $\|x - \bar{x}\| = 8.33$. For comparison, the signal \tilde{x} restored via (1.3) with Algorithm (1.4) is displayed in Fig. 10. In Problem 5.1, this corresponds to $q = 1$, $m = 0$, $\mathbb{K} = \mathbb{N}$, $\tau_{k,l} \equiv 0$, $\omega_k \equiv 2.9$ for the detail coefficients, and $\omega_k \equiv 0.0062$ for the approximation coefficients. This setup yields a worse error of $\|\tilde{x} - \bar{x}\| = 14.14$. The above results have been obtained with a discrete implementation of the wavelet decomposition over 4 resolution levels using 2048 signal samples [27].

References

- [1] A. Antoniadis, D. Leporini, and J.-C. Pesquet, Wavelet thresholding for some classes of non-Gaussian noise, *Statist. Neerlandica*, vol. 56, pp. 434–453, 2002.
- [2] S. Bacchelli and S. Papi, Filtered wavelet thresholding methods, *J. Comput. Appl. Math.*, vol. 164/165, pp. 39–52, 2004.
- [3] H. H. Bauschke, J. V. Burke, F. R. Deutsch, H. S. Hundal, and J. D. Vanderwerff, A new proximal point iteration that converges weakly but not in norm, *Proc. Amer. Math. Soc.*, vol. 133, pp. 1829–1835, 2005.
- [4] H. H. Bauschke, E. Matoušková, and S. Reich, Projection and proximal point methods: Convergence results and counterexamples, *Nonlinear Anal.*, vol. 56, pp. 715–738, 2004.
- [5] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, A ℓ^1 unified variational framework for image restoration, in *Proc. Eighth Europ. Conf. Comput. Vision*, Prague, 2004, T. Pajdla and J. Matas, eds., Lecture Notes in Comput. Sci. 3024, Springer-Verlag, New York, 2004, pp 1–13.
- [6] R. E. Bruck and S. Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston J. Math.*, vol. 3, pp. 459–470, 1977.
- [7] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Process.*, vol. 7, pp. 319–335, 1998.

- [8] S. Chen, D. Donoho, and M. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.*, vol. 43, pp. 129–159, 2001.
- [9] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Trans. Signal Process.*, vol. 42, pp. 2955–2966, 1994.
- [10] P. L. Combettes, A block-iterative surrogate constraint splitting method for quadratic signal recovery, *IEEE Trans. Signal Processing*, vol. 51, pp. 1771–1782, 2003.
- [11] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, vol. 53, pp. 475–504, 2004.
- [12] P. L. Combettes and H. J. Trussell, Method of successive projections for finding a common point of sets in metric spaces, *J. Optim. Theory Appl.*, vol. 67, pp. 487–507, 1990.
- [13] P. L. Combettes and H. J. Trussell, The use of noise properties in set theoretic estimation, *IEEE Trans. Signal Process.*, vol. 39, pp. 1630–1641, 1991.
- [14] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.
- [15] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.
- [16] I. Daubechies, M. Defrise, and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.
- [17] I. Daubechies and G. Teschke, Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising, *Appl. Comput. Harmon. Anal.*, vol. 19, pp. 1–16, 2005.
- [18] C. de Mol and M. Defrise, A note on wavelet-based inversion algorithms, *Contemp. Math.*, vol. 313, pp. 85–96, 2002.
- [19] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, vol. 81, pp. 425–455, 1994.
- [20] D. L. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Stat. Assoc.*, vol. 90, pp. 1200–1224, 1995.
- [21] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, Wavelet shrinkage: Asymptopia?, *J. R. Statist. Soc. B.*, vol. 57, pp. 301–369, 1995.
- [22] M. A. T. Figueiredo and R. D. Nowak, An EM algorithm for wavelet-based image restoration, *IEEE Trans. Image Process.*, vol. 12, pp. 906–916, 2003.
- [23] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.*, vol. 29, pp. 403–419, 1991.
- [24] O. Güler, Convergence rate estimates for the gradient differential inclusion, *Optim. Methods Softw.*, vol. 20, pp. 729–735, 2005.
- [25] P. J. Huber, Robust regression: Asymptotics, conjectures, and Monte Carlo, *Ann. Statist.*, vol. 1, pp. 799–821, 1973.
- [26] T. Kotzer, N. Cohen, and J. Shamir, A projection-based algorithm for consistent and inconsistent constraints, *SIAM J. Optim.*, vol. 7, pp. 527–546, 1997.
- [27] S. G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed, Academic Press, New York, 1999.

- [28] J.-J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.*, vol. 255, pp. 2897–2899, 1962.
- [29] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.
- [30] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.
- [31] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs, *SIAM J. Numer. Anal.*, vol. 42, pp. 686–713, 2004.
- [32] T. Tao and B. Vidakovic, Almost everywhere behavior of general wavelet shrinkage operators, *Appl. Comput. Harmon. Anal.*, vol. 9, pp. 72–82, 2000.
- [33] V. N. Temlyakov, Universal bases and greedy algorithms for anisotropic function classes, *Constr. Approx.*, vol. 18, pp. 529–550, 2002.
- [34] J. A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inform. Theory*, vol. 52, pp. 1030–1051, 2006.
- [35] H. J. Trussell and M. R. Civanlar, The feasible solution in signal restoration, *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, pp. 201–212, 1984.
- [36] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.*, vol. 29, pp. 119–138, 1991.
- [37] B. Vidakovic, Nonlinear wavelet shrinkage with Bayes rules and Bayes factors, *J. Amer. Statist. Assoc.*, vol. 93, pp. 173–179, 1998.
- [38] *WaveLab Toolbox*, Stanford University, <http://www-stat.stanford.edu/~wavelab/>.
- [39] C. Zălinescu, *Convex Analysis in General Vector Spaces*. World Scientific, River Edge, NJ, 2002.