A VARIATIONAL FORMULATION FOR FRAME-BASED INVERSE PROBLEMS*

Caroline Chaux,[1] Patrick L. Combettes,[2] Jean-Christophe Pesquet,[1]
and Valérie R. Wajs[2]

[1]Institut Gaspard Monge and UMR CNRS 8049
Université de Marne la Vallée
77454 Marne la Vallée Cedex 2, France

[2]Laboratoire Jacques-Louis Lions – UMR CNRS 7598
Université Pierre et Marie Curie – Paris 6
75005 Paris, France

Abstract

A convex variational framework is proposed for solving inverse problems in Hilbert spaces with a priori information on the representation of the target solution in a frame. The objective function to be minimized consists of a separable term penalizing each frame coefficient individually and of a smooth term modeling the data formation model as well as other constraints. Sparsity-constrained and Bayesian formulations are examined as special cases. A splitting algorithm is presented to solve this problem and its convergence is established in infinite-dimensional spaces under mild conditions on the penalization functions, which need not be differentiable. Numerical simulations demonstrate applications to frame-based image restoration.

1 Introduction

In inverse problems, certain physical properties of the target solution $\overline{x}$ are most suitably expressed in terms of the coefficients $(\overline{\xi}_k)_{k\in\mathbb{K}\subset\mathbb{N}}$ of its representation $\overline{x} = \sum_{k\in\mathbb{K}} \overline{\xi}_k e_k$ with respect to a family of vectors $(e_k)_{k\in\mathbb{K}}$ in a Hilbert space $(\mathcal{H}, \|\cdot\|)$. Traditionally, such linear representations have been mostly centered on orthonormal bases as, for instance, in Fourier, wavelet, or bandlet decompositions [8, 25, 26]. Recently, attention has shifted towards more general, overcomplete representations known as *frames*; see [6, 7, 16, 20, 33] for specific examples. Recall that a family

*Contact author: P. L. Combettes, plc@math.jussieu.fr, phone: +33 1 4427 6319, fax: +33 1 4427 7200.

of vectors $(e_k)_{k\in\mathbb{K}}$ in $\mathcal{H}$ constitutes a frame if there exist two constants μ and ν in $]0,+\infty[$ such that

$$(\forall x\in\mathcal{H})\ \ \mu\|x\|^2\leq\sum_{k\in\mathbb{K}}|\langle x\mid e_k\rangle|^2\leq\nu\|x\|^2. \tag{1.1}$$

The associated frame operator is the injective bounded linear operator

$$F\colon\mathcal{H}\to\ell^2(\mathbb{K})\colon x\mapsto(\langle x\mid e_k\rangle)_{k\in\mathbb{K}}, \tag{1.2}$$

the adjoint of which is the surjective bounded linear operator

$$F^*\colon\ell^2(\mathbb{K})\to\mathcal{H}\colon(\xi_k)_{k\in\mathbb{K}}\mapsto\sum_{k\in\mathbb{K}}\xi_k e_k. \tag{1.3}$$

When $\mu=\nu$ in (1.1), $(e_k)_{k\in\mathbb{K}}$ is said to be a tight frame. A simple example of a tight frame is the union of m orthonormal bases, in which case $\mu=\nu=m$. For instance, in $\mathcal{H}=\mathrm{L}^2(\mathbb{R}^2)$, a real dual-tree wavelet decomposition is the union of two orthonormal wavelet bases [7, 29]. Curvelets [6] constitute another example of a tight frame of $\mathrm{L}^2(\mathbb{R}^2)$. Historically, Gabor frames [16, 33] have played an important role in many inverse problems. Another common example of a frame is a Riesz basis, which corresponds to the case when $(e_k)_{k\in\mathbb{K}}$ is linearly independent or, equivalently, when F is bijective. In such instances, there exists a unique Riesz basis $(\breve{e}_k)_{k\in\mathbb{K}}$ such that $(e_k)_{k\in\mathbb{K}}$ and $(\breve{e}_k)_{k\in\mathbb{K}}$ are biorthogonal. Furthermore, for every $x\in\mathcal{H}$ and $(\xi_k)_{k\in\mathbb{K}}\in\ell^2(\mathbb{K})$,

$$x=F^*(\xi_k)_{k\in\mathbb{K}}\quad\Leftrightarrow\quad(\forall k\in\mathbb{K})\quad\xi_k=\langle x\mid\breve{e}_k\rangle. \tag{1.4}$$

When $F^{-1}=F^*$, $(e_k)_{k\in\mathbb{K}}$ is an orthonormal basis and $(\breve{e}_k)_{k\in\mathbb{K}}=(e_k)_{k\in\mathbb{K}}$. Examples of Riesz bases of $\mathrm{L}^2(\mathbb{R}^2)$ include biorthogonal bases of compactly supported dyadic wavelets having certain symmetry properties [9]. Further constructions as well as a detailed account of frame theory in Hilbert spaces can be found in [23].

The goal of the present paper is to propose a flexible convex variational framework for solving inverse problems in which a priori information (e.g., sparsity, distribution, statistical properties) is available about the representation of the target solution in a frame. Our formulation (see Problem 2.1) consists of minimizing the sum of a separable, possibly nondifferentiable function penalizing each coefficient of the frame decomposition individually, and of a smooth function which combines other information on the problem and the data formation model. This variational problem is formally stated in Section 2, where connections with sparsity-constrained and Bayesian formulations are established. Section 3 is devoted to proximity operators, which will be an essential tool in our algorithm. In connection with the Bayesian framework discussed in Section 2.3, we derive in Section 4 closed-form expressions for the proximity operators associated with a variety of univariate log-concave distributions. The question of the existence of solutions to Problem 2.1 is addressed in Section 5. An inexact proximal algorithm for solving Problem 2.1 is presented in Section 6 and its convergence is established in infinite-dimensional spaces under mild assumptions on the penalization functions. An attractive feature of this algorithm is that it is fully split in that, at each iteration, all the functions appearing in the problem are activated individually. Finally, applications to image recovery are demonstrated in Section 7.

2 Problem formulation

2.1 Notation, assumptions, and problem statement

Throughout, $\mathcal{H}$ is a separable real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$, norm $\|\cdot\|$, and distance d. $\Gamma_0(\mathcal{H})$ is the class of lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$ which are not identically equal to $+\infty$. The support function of a subset S of $\mathcal{H}$ is $\sigma_S \colon \mathcal{H} \to [-\infty, +\infty] : u \mapsto \sup_{x \in S} \langle x \mid u \rangle$, and its distance function is $d_S \colon \mathcal{H} \to [0, +\infty] : x \mapsto \inf \|S - x\|$. If S is nonempty, closed, and convex then, for every $x \in \mathcal{H}$, there exists a unique point $P_S x$ in S, called the projection of x onto S, such that $\|x - P_S x\| = d_S(x)$ (further background on convex analysis will be found in [35]). The index set $\mathbb{K}$ is either $\{1, \ldots, K\}$ ($K \in \mathbb{N}$) or $\mathbb{N}$, according as $\mathcal{H}$ is finite or infinite dimensional. Moreover, $(e_k)_{k \in \mathbb{K}}$ is a frame in $\mathcal{H}$ with constants μ and ν (see (1.1)) and frame operator F (see (1.2)). Finally, the sequence of frame coefficients of a generic point $x \in \mathcal{H}$ will be denoted by x, i.e., $\mathsf{x} = (\xi_k)_{k \in \mathbb{K}}$, where $x = \sum_{k \in \mathbb{K}} \xi_k e_k$.

Let $\overline{x} \in \mathcal{H}$ be the target solution of the underlying inverse problem. Our basic premise is that a priori information is available about the coefficients $(\overline{\xi}_k)_{k \in \mathbb{K}}$ of the decomposition

$$\overline{x} = \sum_{k \in \mathbb{K}} \overline{\xi}_k e_k \tag{2.1}$$

of $\overline{x}$ in $(e_k)_{k \in \mathbb{K}}$. To recover $\overline{x}$, it is therefore natural to formulate a variational problem in the space $\ell^2(\mathbb{K})$ of frame coefficients, where a priori information on $(\overline{\xi}_k)_{k \in \mathbb{K}}$ can be easily incorporated. More precisely, a solution will assume the form $\tilde{x} = \sum_{k \in \mathbb{K}} \tilde{\xi}_k e_k$, where $(\tilde{\xi}_k)_{k \in \mathbb{K}}$ is a solution to the following problem (see Definition 3.1 for the notation $\operatorname{prox}_{\phi_k}$ in (iii)).

Problem 2.1 Let $(\phi_k)_{k \in \mathbb{K}}$ be functions in $\Gamma_0(\mathbb{R})$ such that either $\mathbb{K} = \{1, \ldots, K\}$ with $K \in \mathbb{N}$, or $\mathbb{K} = \mathbb{N}$ and there exists a subset $\mathbb{L}$ of $\mathbb{K}$ such that

(i) $\mathbb{K} \smallsetminus \mathbb{L}$ is finite;

(ii) $(\forall k \in \mathbb{L})\ \phi_k \geq 0$;

(iii) there exists a sequence $(\zeta_k)_{k \in \mathbb{L}}$ in $\mathbb{R}$ such that $\sum_{k \in \mathbb{L}} |\zeta_k|^2 < +\infty$, $\sum_{k \in \mathbb{L}} |\operatorname{prox}_{\phi_k} \zeta_k|^2 < +\infty$, and $\sum_{k \in \mathbb{L}} \phi_k(\zeta_k) < +\infty$.

In addition, let $\Psi \in \Gamma_0(\mathcal{H})$ be differentiable on $\mathcal{H}$ with a τ-Lipschitz continuous gradient for some $\tau \in]0, +\infty[$. The objective is to

$$\underset{(\xi_k)_{k \in \mathbb{K}} \in \ell^2(\mathbb{K})}{\text{minimize}} \ \sum_{k \in \mathbb{K}} \phi_k(\xi_k) + \Psi\left(\sum_{k \in \mathbb{K}} \xi_k e_k\right). \tag{2.2}$$

Remark 2.2

(i) The functions $(\phi_k)_{k \in \mathbb{K}}$ in Problem 2.1 need not be differentiable. As will be seen in Section 2.2, this feature is essential in sparsity-constrained problems.

(ii) Suppose that $\mathbb{K} = \mathbb{N}$. Then Conditions (ii) and (iii) in Problem 2.1 hold when, for every $k \in \mathbb{L}$, ϕ_k admits a minimizer ζ_k such that $\phi_k(\zeta_k) = 0$ and $\sum_{k\in\mathbb{L}} |\zeta_k|^2 < +\infty$. Indeed, we then have $\sum_{k\in\mathbb{L}} |\operatorname{prox}_{\phi_k} \zeta_k|^2 = \sum_{k\in\mathbb{L}} |\zeta_k|^2 < +\infty$ (this will follow from Lemma 3.3(i)) and $\sum_{k\in\mathbb{L}} \phi_k(\zeta_k) = 0$. In particular, Conditions (ii) and (iii) in Problem 2.1 hold when $(\forall k \in \mathbb{L})$ $\phi_k \geq \phi_k(0) = 0$. This amounts to setting $\zeta_k \equiv 0$ in the previous item.

In Problem 2.1, the functions $(\phi_k)_{k\in\mathbb{K}}$ penalize the frame coefficients $(\xi_k)_{k\in\mathbb{K}}$, while the function Ψ penalizes $x = F^*(\xi_k)_{k\in\mathbb{K}} = \sum_{k\in\mathbb{K}} \xi_k e_k$, thereby modeling direct constraints on $\overline{x}$. This flexible framework makes it possible to model a wide range of inverse problems. Two important instances are presented below.

2.2 Inverse problems with sparsity constraints

A common objective in selecting the frame $(e_k)_{k\in\mathbb{K}}$ is to obtain a sparse representation of the target solution $\overline{x}$ in the sense that most of the coefficients $(\overline{\xi}_k)_{k\in\mathbb{K}}$ in (2.1) are zero. By choosing $\phi_k = \omega_k|\cdot|$ with $\omega_k > 0$ in Problem 2.1, one aims at setting to zero the kth coefficient if it falls in the interval $[-\omega_k, \omega_k]$, hence promoting sparsity (see [17, 22, 32] for special cases). More generally, to aim at zeroing a coefficient falling in a closed interval $\Omega_k \subset \mathbb{R}$, one can use the function $\phi_k = \psi_k + \sigma_{\Omega_k}$, where ψ_k satisfies $0 = \psi_k(0) \leq \psi_k \in \Gamma_0(\mathbb{R})$ and is differentiable at 0 [14]. This construct actually characterizes all thresholders on Ω_k that have properties suitable to their use in iterative methods [14, Theorem 3.3].

Now assume that the problem under consideration is to recover $\overline{x} \in \mathcal{H}$ from q observations

$$z_i = T_i\overline{x} + v_i, \quad 1 \leq i \leq q, \tag{2.3}$$

where T_i is a bounded linear operator from $\mathcal{H}$ to a real Hilbert space $\mathcal{G}_i$, $z_i \in \mathcal{G}_i$, and $v_i \in \mathcal{G}_i$ is the realization of a noise process. A standard data fidelity criterion in such instances is the function $x \mapsto \sum_{i=1}^q \alpha_i\|T_i x - z_i\|^2$, where $(\alpha_i)_{1\leq i\leq q}$ are strictly positive reals, see e.g., [12, 21]. In addition, assume that a priori information is available that constrains $\overline{x}$ to lie in some closed convex subsets $(S_i)_{1\leq i\leq m}$ of $\mathcal{H}$ (see [10, 30] and the references therein for examples). These constraints can be aggregated via the cost function $x \mapsto \sum_{i=1}^m \beta_i d_{S_i}^2(x)$, where $(\beta_i)_{1\leq i\leq m}$ are strictly positive reals [5, 11]. These two objectives can be combined by using the function

$$\Psi\colon x \mapsto \frac{1}{2}\sum_{i=1}^q \alpha_i\|T_i x - z_i\|^2 + \frac{1}{2}\sum_{i=1}^m \beta_i d_{S_i}^2(x) \tag{2.4}$$

in Problem 2.1. This function is indeed differentiable and its gradient

$$\nabla\Psi\colon x \mapsto \sum_{i=1}^q \alpha_i T_i^*(T_i x - z_i) + \sum_{i=1}^m \beta_i(x - P_{S_i}x) \tag{2.5}$$

has Lipschitz constant [14, Section 5.1]

$$\tau = \Big\|\sum_{i=1}^q \alpha_i T_i^* T_i\Big\| + \sum_{i=1}^m \beta_i. \tag{2.6}$$

In instances when $\|\sum_{i=1}^q \alpha_i T_i^* T_i\|$ cannot be evaluated directly, it can be majorized by $\sum_{i=1}^q \alpha_i \|T_i\|^2$. It should be noted that, more generally, Ψ remains Lipschitz-continuous if the term $\sum_{i=1}^m \beta_i d_{S_i}^2(x)$ in (2.4) is replaced by a sum of Moreau envelopes, see [13, Section 6.3] and [15, Section 4.1] for details.

2.3 Bayesian statistical framework

A standard linear inverse problem is to recover $\overline{x} \in \mathcal{H}$ from an observation

$$z = T\overline{x} + v, \tag{2.7}$$

in a real Hilbert space $\mathcal{G}$, where $T \colon \mathcal{H} \to \mathcal{G}$ is a bounded linear operator and where $v \in \mathcal{G}$ stands for an additive noise perturbation. If $\overline{\mathsf{x}} = (\overline{\xi}_k)_{k\in\mathbb{K}}$ denotes the coefficients of $\overline{x}$ in $(e_k)_{k\in\mathbb{K}}$, (2.7) can be written as

$$z = TF^*\overline{\mathsf{x}} + v. \tag{2.8}$$

For the sake of simplicity, the following assumptions regarding (2.8) are made in this section (with the usual convention $\ln 0 = -\infty$).

Assumption 2.3

(i) $\mathcal{H} = \mathbb{R}^N$, $\mathcal{G} = \mathbb{R}^M$, and $\mathbb{K} = \{1, \ldots, K\}$, where $K \geq N$.

(ii) The vectors $\overline{\mathsf{x}}$, z, and v are, respectively, realizations of real-valued random vectors $\overline{\mathsf{X}}$, Z, and V defined on the same probability space.

(iii) The random vectors $\overline{\mathsf{X}}$ and V are mutually independent and have probability density functions $f_{\overline{\mathsf{X}}}$ and f_V, respectively.

(iv) The components of $\overline{\mathsf{X}}$ are independent with upper-semicontinuous log-concave densities.

(v) The function $\ln f_V$ is concave and differentiable with a Lipschitz continuous gradient.

Under Assumption 2.3, a common Bayesian approach for estimating $\overline{\mathsf{x}}$ from z consists in applying a maximum a posteriori (MAP) rule [3, 4, 31], which amounts to maximizing the posterior probability density $f_{\overline{\mathsf{X}}|Z=z}$. Thus, $\tilde{\mathsf{x}}$ is a MAP estimate of $\overline{\mathsf{x}}$ if

$$(\forall \mathsf{x} \in \mathbb{R}^K) \qquad f_{\overline{\mathsf{X}}|Z=z}(\tilde{\mathsf{x}}) \geq f_{\overline{\mathsf{X}}|Z=z}(\mathsf{x}). \tag{2.9}$$

Using Bayes' formula, this amounts to solving

$$\underset{\mathsf{x}\in\mathbb{R}^K}{\text{minimize}} \quad -\ln f_{\overline{\mathsf{X}}}(\mathsf{x}) - \ln f_{Z|\overline{\mathsf{X}}=\mathsf{x}}(z). \tag{2.10}$$

In view of (2.8), this is also equivalent to solving

$$\underset{\mathsf{x}\in\mathbb{R}^K}{\text{minimize}} \quad -\ln f_{\overline{\mathsf{X}}}(\mathsf{x}) - \ln f_V(z - TF^*\mathsf{x}). \tag{2.11}$$

Under Assumption 2.3, this convex optimization problem is a special case of Problem 2.1. Indeed, Assumption 2.3(iv) allows us to write, without loss of generality, the prior density as

$$(\forall (\xi_k)_{k\in\mathbb{K}} \in \mathbb{R}^K) \quad f_{\overline{\mathsf{X}}}((\xi_k)_{k\in\mathbb{K}}) \propto \prod_{k=1}^{K} \exp(-\phi_k(\xi_k)), \tag{2.12}$$

where $(\phi_k)_{k\in\mathbb{K}}$ are the so-called potential functions of the marginal probability density functions of $\overline{\mathsf{X}}$. It also follows from Assumption 2.3(iv) that the functions $(\phi_k)_{k\in\mathbb{K}}$ are in $\Gamma_0(\mathbb{R})$. Now set

$$(\forall x \in \mathcal{H}) \quad \Psi(x) = -\ln f_V(z - Tx). \tag{2.13}$$

Then Assumption 2.3(v) asserts that $\Psi \in \Gamma_0(\mathcal{H})$ is differentiable with a Lipschitz continuous gradient. Altogether, (2.11) reduces to Problem 2.1.

Remark 2.4 In the simple case when V is a zero-mean Gaussian vector with an invertible covariance matrix Λ, the function Ψ reduces (up to an additive constant) to the residual energy function $x \mapsto \langle \Lambda^{-1}(z - Tx) \mid z - Tx\rangle/2$. When $\overline{\mathsf{X}}$ is further assumed to be Gaussian, the solution to Problem 2.1 is a linear function of z. Recall that the MAP estimate coincides with the minimum mean-square error estimate under such Gaussian models for both V and $\overline{\mathsf{X}}$ [34, Section 2.4].

Remark 2.5 An alternative Bayesian strategy would be to determine a MAP estimate of $\overline{x}$. This would lead to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad -\ln f_{\overline{X}}(x) - \ln f_V(z - Tx), \tag{2.14}$$

where $f_{\overline{X}}$ can be deduced from (2.12) through the change of variable $\overline{X} = F^*\overline{\mathsf{X}}$. In the case of an orthonormal basis decomposition, it is easy to check that (2.14) is equivalent to problem (2.11). By contrast, when F corresponds to an overcomplete frame, the expression of $f_{\overline{X}}$ becomes involved and (2.14) is usually much less tractable than Problem 2.1. As will be seen in Section 6, the latter can be solved via a simple splitting algorithm.

Remark 2.6 Let us decompose the observation vector as $z = [z_1^\top, \dots, z_q^\top]^\top$ and the matrix representing T as $[T_1^\top, \dots, T_q^\top]^\top$ where, for every $i \in \{1, \dots, q\}$, $z_i \in \mathbb{R}^{M_i}$ and $T_i \in \mathbb{R}^{M_i\times N}$ with $\sum_{i=1}^q M_i = M$. Furthermore, assume that V is a zero-mean Gaussian vector with diagonal covariance matrix

$$\Lambda = \begin{bmatrix} \alpha_1^{-1} I_{M_1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \alpha_q^{-1} I_{M_q} \end{bmatrix}, \tag{2.15}$$

where $(\alpha_i)_{1\le i\le q}$ are strictly positive reals and I_{M_i}, $1 \le i \le q$, is the identity matrix of size $M_i \times M_i$. Then Ψ reduces to the first term in (2.4) where $\mathcal{G}_i = \mathbb{R}^{M_i}$ and the MAP estimation problem under Assumption 2.3 becomes a special case of the problem addressed in Section 2.2 with $m = 0$.

3 Basic tool: proximity operator

In this section, $\mathcal{X}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$ and norm $\|\cdot\|$.

3.1 Background

Let $\varphi \in \Gamma_0(\mathcal{X})$. The subdifferential of φ at $\mathsf{x} \in \mathcal{X}$ is the set

$$\partial\varphi(\mathsf{x}) = \big\{\mathsf{u} \in \mathcal{X} \;\big|\; (\forall \mathsf{y} \in \mathcal{X})\; \langle \mathsf{y} - \mathsf{x} \mid \mathsf{u}\rangle + \varphi(\mathsf{x}) \leq \varphi(\mathsf{y})\big\}. \tag{3.1}$$

If φ is Gâteaux differentiable at x with gradient $\nabla\varphi(\mathsf{x})$, then $\partial\varphi(\mathsf{x}) = \{\nabla\varphi(\mathsf{x})\}$. The conjugate of φ is the function $\varphi^* \in \Gamma_0(\mathcal{X})$ defined by

$$(\forall \mathsf{u} \in \mathcal{X}) \quad \varphi^*(\mathsf{u}) = \sup_{\mathsf{x} \in \mathcal{X}} \langle \mathsf{x} \mid \mathsf{u}\rangle - \varphi(\mathsf{x}). \tag{3.2}$$

In connection with the analysis and the numerical solution of Problem 2.1, an extremely useful tool is the following.

Definition 3.1 [27] Let $\varphi \in \Gamma_0(\mathcal{X})$. Then, for every $\mathsf{x} \in \mathcal{X}$, the function $\mathsf{y} \mapsto \varphi(\mathsf{y}) + \|\mathsf{x} - \mathsf{y}\|^2/2$ achieves its infimum at a unique point denoted by $\operatorname{prox}_\varphi \mathsf{x}$. The operator $\operatorname{prox}_\varphi \colon \mathcal{X} \to \mathcal{X}$ thus defined is the *proximity operator* of φ. Moreover,

$$(\forall \mathsf{x} \in \mathcal{X})(\forall \mathsf{p} \in \mathcal{X}) \quad \mathsf{p} = \operatorname{prox}_\varphi \mathsf{x} \;\Leftrightarrow\; \mathsf{x} - \mathsf{p} \in \partial\varphi(\mathsf{p}) \tag{3.3}$$

$$\Leftrightarrow\; (\forall \mathsf{y} \in \mathcal{X})\;\; \langle \mathsf{y} - \mathsf{p} \mid \mathsf{x} - \mathsf{p}\rangle + \varphi(\mathsf{p}) \leq \varphi(\mathsf{y}). \tag{3.4}$$

Here are basic properties of this operator.

Lemma 3.2 [15, Lemma 2.6] *Let $\varphi \in \Gamma_0(\mathcal{X})$ and $\mathsf{x} \in \mathcal{X}$. Then the following hold.*

(i) *Let $\psi = \varphi + \alpha\|\cdot\|^2/2 + \langle \cdot \mid \mathsf{u}\rangle + \beta$, where $\mathsf{u} \in \mathcal{X}$, $\alpha \in [0, +\infty[$, and $\beta \in \mathbb{R}$. Then $\operatorname{prox}_\psi \mathsf{x} = \operatorname{prox}_{\varphi/(\alpha+1)} \big((\mathsf{x} - \mathsf{u})/(\alpha + 1)\big)$.*

(ii) *Let $\psi = \varphi(\cdot - \mathsf{z})$, where $\mathsf{z} \in \mathcal{X}$. Then $\operatorname{prox}_\psi \mathsf{x} = \mathsf{z} + \operatorname{prox}_\varphi(\mathsf{x} - \mathsf{z})$.*

(iii) *Let $\psi = \varphi(\cdot/\rho)$, where $\rho \in \mathbb{R} \smallsetminus \{0\}$. Then $\operatorname{prox}_\psi \mathsf{x} = \rho \operatorname{prox}_{\varphi/\rho^2}(\mathsf{x}/\rho)$.*

(iv) *Let $\psi \colon \mathsf{y} \mapsto \varphi(-\mathsf{y})$. Then $\operatorname{prox}_\psi \mathsf{x} = -\operatorname{prox}_\varphi(-\mathsf{x})$.*

Lemma 3.3 [15, Section 2] *Let $\varphi \in \Gamma_0(\mathcal{X})$. Then the following hold.*

(i) $(\forall \mathsf{x} \in \mathcal{X})\; \mathsf{x} \in \operatorname{Argmin} \varphi \Leftrightarrow 0 \in \partial\varphi(\mathsf{x}) \Leftrightarrow \operatorname{prox}_\varphi \mathsf{x} = \mathsf{x}$.

(ii) $(\forall \mathsf{x} \in \mathcal{X})(\forall \mathsf{y} \in \mathcal{X})\; \|\operatorname{prox}_\varphi \mathsf{x} - \operatorname{prox}_\varphi \mathsf{y}\| \leq \|\mathsf{x} - \mathsf{y}\|$.

(iii) $(\forall \mathsf{x} \in \mathcal{X})(\forall \gamma \in \,]0, +\infty[)\; \mathsf{x} = \operatorname{prox}_{\gamma\varphi} \mathsf{x} + \gamma \operatorname{prox}_{\varphi^*/\gamma}(\mathsf{x}/\gamma)$.

3.2 Forward-backward splitting

In this section, we consider the following abstract variational framework for Problem 2.1.

Problem 3.4 Let f_1 and f_2 be functions in $\Gamma_0(\mathcal{X})$ such that f_2 is differentiable on $\mathcal{X}$ with a β-Lipschitz continuous gradient for some $\beta \in \,]0, +\infty[$. The objective is to

$$\underset{\mathsf{x}\in\mathcal{X}}{\text{minimize}} \;\; f_1(\mathsf{x}) + f_2(\mathsf{x}). \tag{3.5}$$

The following result describes a proximal algorithm for solving this problem which is based on the forward-backward splitting method for monotone operators [13].

Theorem 3.5 [15, Theorem 3.4(i)] *Suppose that* $\operatorname{Argmin}(f_1 + f_2) \neq \varnothing$. *Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0, +\infty[$ *such that* $0 < \inf_{n\in\mathbb{N}} \gamma_n \leq \sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0, 1]$ *such that* $\inf_{n\in\mathbb{N}} \lambda_n > 0$, *and let* $(\mathsf{a}_n)_{n\in\mathbb{N}}$ *and* $(\mathsf{b}_n)_{n\in\mathbb{N}}$ *be sequences in* $\mathcal{X}$ *such that* $\sum_{n\in\mathbb{N}} \|\mathsf{a}_n\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|\mathsf{b}_n\| < +\infty$. *Fix* $\mathsf{x}_0 \in \mathcal{X}$ *and, for every* $n \in \mathbb{N}$, *set*

$$\mathsf{x}_{n+1} = \mathsf{x}_n + \lambda_n \Big(\operatorname{prox}_{\gamma_n f_1} \big(\mathsf{x}_n - \gamma_n(\nabla f_2(\mathsf{x}_n) + \mathsf{b}_n)\big) + \mathsf{a}_n - \mathsf{x}_n \Big). \tag{3.6}$$

Then $(\mathsf{x}_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 3.4.*

3.3 Decomposition formula

The following decomposition property will be instrumental in our analysis.

Proposition 3.6 *Set* $\Upsilon\colon \mathcal{X} \to \,]-\infty, +\infty] : \mathsf{x} \mapsto \sum_{i\in\mathbb{I}} \psi_i(\langle \mathsf{x} \mid \mathsf{o}_i\rangle)$, *where:*

(i) $\varnothing \neq \mathbb{I} \subset \mathbb{N}$;

(ii) $(\mathsf{o}_i)_{i\in\mathbb{I}}$ *is an orthonormal basis of* $\mathcal{X}$;

(iii) $(\psi_i)_{i\in\mathbb{I}}$ *are functions in* $\Gamma_0(\mathbb{R})$;

(iv) *Either* $\mathbb{I}$ *is finite, or there exists a subset* $\mathbb{J}$ *of* $\mathbb{I}$ *such that:*

 (a) $\mathbb{I} \smallsetminus \mathbb{J}$ *is finite;*

 (b) $(\forall i \in \mathbb{J})\ \psi_i \geq 0$;

 (c) *there exists a sequence* $(\zeta_i)_{i\in\mathbb{J}}$ *in* $\mathbb{R}$ *such that* $\sum_{i\in\mathbb{J}} |\zeta_i|^2 < +\infty$, $\sum_{i\in\mathbb{J}} |\operatorname{prox}_{\psi_i} \zeta_i|^2 < +\infty$, *and* $\sum_{i\in\mathbb{J}} \psi_i(\zeta_i) < +\infty$.

Then $\Upsilon \in \Gamma_0(\mathcal{X})$ *and* $(\forall \mathsf{x} \in \mathcal{X})\ \operatorname{prox}_\Upsilon \mathsf{x} = \sum_{i\in\mathbb{I}} \big(\operatorname{prox}_{\psi_i} \langle \mathsf{x} \mid \mathsf{o}_i\rangle\big)\mathsf{o}_i$.

Proof. We treat only the case when $\mathbb{I}$ is infinite as the case when $\mathbb{I}$ is finite will follow trivially the arguments presented below. Fix, for every $i \in \mathbb{I} \smallsetminus \mathbb{J}$, $\zeta_i \in \mathbb{R}$ such that $\psi_i(\zeta_i) < +\infty$ and set $\mathsf{z} = \sum_{i\in\mathbb{I}} \zeta_i \mathsf{o}_i$. Then (iv) implies that $\sum_{i\in\mathbb{I}} \zeta_i^2 < +\infty$ and, in view of (ii), that $\mathsf{z} \in \mathcal{X}$. Moreover, $\Upsilon(\mathsf{z}) = \sum_{i\in\mathbb{J}} \psi_i(\zeta_i) + \sum_{i\in\mathbb{I}\smallsetminus\mathbb{J}} \psi_i(\zeta_i) < +\infty$.

Let us show that $\Upsilon \in \Gamma_0(\mathcal{X})$. As just seen, $\Upsilon(\mathsf{z}) < +\infty$ and, therefore, $\Upsilon \not\equiv +\infty$. Next, we observe that, by virtue of (iii), the functions $(\psi_i(\langle \cdot \mid \mathsf{o}_i\rangle))_{i\in\mathbb{I}}$ are lower semicontinuous and convex. As a result, $\sum_{i\in\mathbb{I}\smallsetminus\mathbb{J}} \psi_i(\langle \cdot \mid \mathsf{o}_i\rangle)$ is lower semicontinuous and convex, as a finite sum of such functions. Thus, to show that $\Upsilon \in \Gamma_0(\mathcal{X})$, it remains to show that $\Upsilon_{\mathbb{J}} = \sum_{i\in\mathbb{J}} \psi_i(\langle \cdot \mid \mathsf{o}_i\rangle)$ is lower semicontinuous and convex. It follows from (iv)(b) that

$$\Upsilon_{\mathbb{J}} = \sup_{\substack{\mathbb{J}'\subset\mathbb{J}\\ \mathbb{J}' \text{ finite}}} \sum_{i\in\mathbb{J}'} \psi_i(\langle \cdot \mid \mathsf{o}_i\rangle). \tag{3.7}$$

However, as above, each finite sum $\sum_{i\in\mathbb{J}'} \psi_i(\langle \cdot \mid \mathsf{o}_i\rangle)$ is lower semicontinuous and convex. Therefore, $\Upsilon_{\mathbb{J}}$ is likewise as the supremum of a family of lower semicontinuous convex functions.

Now fix $\mathsf{x} \in \mathcal{X}$ and set

$$(\forall i \in \mathbb{I}) \quad \xi_i = \langle \mathsf{x} \mid \mathsf{o}_i\rangle \quad \text{and} \quad \pi_i = \operatorname{prox}_{\psi_i} \xi_i. \tag{3.8}$$

It follows from (iv)(a) and (iv)(c) that

$$\sum_{i\in\mathbb{I}} |\operatorname{prox}_{\psi_i} \zeta_i|^2 = \sum_{i\in\mathbb{J}} |\operatorname{prox}_{\psi_i} \zeta_i|^2 + \sum_{i\in\mathbb{I}\smallsetminus\mathbb{J}} |\operatorname{prox}_{\psi_i} \zeta_i|^2 < +\infty. \tag{3.9}$$

Hence, we derive from Lemma 3.3(ii) and (ii) that

$$\begin{aligned}\frac{1}{2}\sum_{i\in\mathbb{I}} |\pi_i|^2 &\leq \sum_{i\in\mathbb{I}} |\pi_i - \operatorname{prox}_{\psi_i} \zeta_i|^2 + \sum_{i\in\mathbb{I}} |\operatorname{prox}_{\psi_i} \zeta_i|^2\\ &= \sum_{i\in\mathbb{I}} |\operatorname{prox}_{\psi_i} \xi_i - \operatorname{prox}_{\psi_i} \zeta_i|^2 + \sum_{i\in\mathbb{I}} |\operatorname{prox}_{\psi_i} \zeta_i|^2\\ &\leq \sum_{i\in\mathbb{I}} |\xi_i - \zeta_i|^2 + \sum_{i\in\mathbb{I}} |\operatorname{prox}_{\psi_i} \zeta_i|^2\\ &= \|\mathsf{x} - \mathsf{z}\|^2 + \sum_{i\in\mathbb{I}} |\operatorname{prox}_{\psi_i} \zeta_i|^2\\ &< +\infty.\end{aligned} \tag{3.10}$$

Let us set $\mathsf{p} = \sum_{i\in\mathbb{I}} \pi_i \mathsf{o}_i$. Then it follows from (3.10) and (ii) that $\mathsf{p} \in \mathcal{X}$. On the other hand, we derive from (3.8) and (3.4) that

$$(\forall i \in \mathbb{I})(\forall \eta \in \mathbb{R}) \quad (\eta - \pi_i)(\xi_i - \pi_i) + \psi_i(\pi_i) \leq \psi_i(\eta). \tag{3.11}$$

Hence, by Parseval and (ii),

$$\begin{aligned}(\forall \mathsf{y} \in \mathcal{X}) \quad \langle \mathsf{y}-\mathsf{p} \mid \mathsf{x}-\mathsf{p}\rangle + \Upsilon(\mathsf{p}) &= \sum_{i\in\mathbb{I}} \langle \mathsf{y}-\mathsf{p} \mid \mathsf{o}_i\rangle\langle \mathsf{x}-\mathsf{p} \mid \mathsf{o}_i\rangle + \sum_{i\in\mathbb{I}} \psi_i(\pi_i)\\ &= \sum_{i\in\mathbb{I}} (\langle \mathsf{y} \mid \mathsf{o}_i\rangle - \pi_i)(\xi_i-\pi_i) + \psi_i(\pi_i)\\ &\le \sum_{i\in\mathbb{I}} \psi_i(\langle \mathsf{y} \mid \mathsf{o}_i\rangle)\\ &= \Upsilon(\mathsf{y}).\end{aligned} \tag{3.12}$$

Invoking (3.4) once again, we conclude that $\mathsf{p} = \operatorname{prox}_\Upsilon \mathsf{x}$. ☐

4 Proximity operators associated with log-concave densities

As discussed in Section 2.3, the functions $(\phi_k)_{k\in\mathbb{K}}$ in (2.2) act as the potential functions of log-concave univariate probability densities modeling the frame coefficients individually in Bayesian formulations. On the other hand, the proximity operators of such functions will, via Proposition 3.6, play a central role in Section 6. Hereafter, we derive closed-form expressions for these proximity operators in the case of some classical log-concave univariate probability densities [19, Chapters VII&IX]. We start with some technical facts.

Lemma 4.1 [14, Section 2] *Let $\phi \in \Gamma_0(\mathbb{R})$. Then the following hold.*

(i) $\operatorname{prox}_\phi \colon \mathbb{R} \to \mathbb{R}$ *is nondecreasing.*

(ii) *Suppose that ϕ admits 0 as a minimizer. Then*

$$(\forall \xi \in \mathbb{R}) \quad \begin{cases} 0 \le \operatorname{prox}_\phi \xi \le \xi, & \text{if } \xi > 0;\\ \operatorname{prox}_\phi \xi = 0, & \text{if } \xi = 0;\\ \xi \le \operatorname{prox}_\phi \xi \le 0, & \text{if } \xi < 0.\end{cases} \tag{4.1}$$

This is true in particular when ϕ is even.

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

$$\operatorname{soft}_\Omega : \mathbb{R} \to \mathbb{R} \colon \xi \mapsto \begin{cases} \xi - \underline{\omega}, & \text{if } \xi < \underline{\omega};\\ 0, & \text{if } \xi \in \Omega;\\ \xi - \overline{\omega}, & \text{if } \xi > \overline{\omega},\end{cases} \qquad \text{with} \quad \begin{cases} \underline{\omega} = \inf \Omega,\\ \overline{\omega} = \sup \Omega,\end{cases} \tag{4.2}$$

is the soft thresholder on Ω.

Remark 4.2 Let $\phi \in \Gamma_0(\mathbb{R})$.

(i) It follows from Definition 3.1 that $(\forall \xi \in \mathbb{R})\ \phi(\operatorname{prox}_\phi \xi) < +\infty$.

(ii) If ϕ is even, then it follows from Lemma 3.2(iv) that prox_ϕ is odd. Therefore, in such instances, it will be enough to determine $\operatorname{prox}_\phi \xi$ for $\xi \geq 0$ and to extend the result to $\xi < 0$ by antisymmetry.

(iii) Let $\xi \in \mathbb{R}$. If ϕ is differentiable at $\operatorname{prox}_\phi \xi$, then (3.3) yields

$$(\forall \pi \in \mathbb{R}) \quad \pi = \operatorname{prox}_\phi \xi \quad \Leftrightarrow \quad \pi + \phi'(\pi) = \xi. \tag{4.3}$$

Let us now examine some concrete examples.

Example 4.3 (Laplace distribution) Let $\omega \in \,]0, +\infty[$ and set

$$\phi\colon \mathbb{R} \to \,]-\infty, +\infty] : \xi \mapsto \omega|\xi|. \tag{4.4}$$

Then, for every $\xi \in \mathbb{R}$, $\operatorname{prox}_\phi \xi = \operatorname{soft}_{[-\omega,\omega]} \xi = \operatorname{sign}(\xi)\max\{|\xi| - \omega, 0\}$.

Proof. Apply Lemma 4.1(iii) with $\psi = 0$ and $\Omega = [-\omega, \omega]$. ☐

Example 4.4 (Gaussian distribution) Let $\tau \in \,]0, +\infty[$ and set

$$\phi\colon \mathbb{R} \to \,]-\infty, +\infty] : \xi \mapsto \tau|\xi|^2. \tag{4.5}$$

Then, for every $\xi \in \mathbb{R}$, $\operatorname{prox}_\phi \xi = \xi/(2\tau + 1)$.

Proof. Apply Lemma 3.2(i) with $\mathcal{X} = \mathbb{R}$, $\varphi = 0$, $\alpha = 2\tau$, and $\mathsf{u} = 0$. ☐

Example 4.5 (generalized Gaussian distribution) Let $p \in \,]1, +\infty[$, $\kappa \in \,]0, +\infty[$, and set

$$\phi\colon \mathbb{R} \to \,]-\infty, +\infty] : \xi \mapsto \kappa|\xi|^p. \tag{4.6}$$

Then, for every $\xi \in \mathbb{R}$, $\operatorname{prox}_\phi \xi = \operatorname{sign}(\xi)\varrho$ where ϱ is the unique nonnegative solution to

$$\varrho + p\kappa\varrho^{p-1} = |\xi|. \tag{4.7}$$

In particular, the following hold:

(i) $\operatorname{prox}_\phi \xi = \xi + \dfrac{4\kappa}{3 \cdot 2^{1/3}}\Big((\chi - \xi)^{1/3} - (\chi + \xi)^{1/3}\Big)$, where $\chi = \sqrt{\xi^2 + 256\kappa^3/729}$, if $p = 4/3$;

(ii) $\operatorname{prox}_\phi \xi = \xi + 9\kappa^2 \operatorname{sign}(\xi)\big(1 - \sqrt{1 + 16|\xi|/(9\kappa^2)}\,\big)/8$, if $p = 3/2$;

(iii) $\operatorname{prox}_\phi \xi = \operatorname{sign}(\xi)\big(\sqrt{1 + 12\kappa|\xi|} - 1\big)/(6\kappa)$, if $p = 3$;

(iv) $\operatorname{prox}_\phi \xi = \left(\dfrac{\chi + \xi}{8\kappa}\right)^{1/3} - \left(\dfrac{\chi - \xi}{8\kappa}\right)^{1/3}$, where $\chi = \sqrt{\xi^2 + 1/(27\kappa)}$, if $p = 4$.

Proof. Let $\xi \in \mathbb{R}$ and set $\pi = \mathrm{prox}_\phi \xi$. As seen in Remark 4.2(ii), because ϕ is even, it is enough to assume that $\xi \geq 0$. Since ϕ is differentiable, it follows from (4.1) and (4.3) that π is the unique nonnegative solution to

$$\pi + p\kappa\pi^{p-1} = \xi, \tag{4.8}$$

which provides (4.7). For $p = 3$, π is the nonnegative solution to the equation $\pi + 3\kappa\pi^2 - \xi = 0$, i.e., $\pi = (\sqrt{1 + 12\kappa\xi} - 1)/(6\kappa)$ and we obtain (iii) by antisymmetry. In turn, since $\big(2|\cdot|^{3/2}/3\big)^* = |\cdot|^3/3$, Lemma 3.3(iii) with $\gamma = 3\kappa/2$ yields $\pi = \mathrm{prox}_{\gamma(2|\cdot|^{3/2}/3)} \xi = \xi - \gamma\, \mathrm{prox}_{(3\gamma)^{-1}|\cdot|^3}(\xi/\gamma) = \xi + 9\kappa^2 \,\mathrm{sign}(\xi)\big(1 - \sqrt{1 + 16|\xi|/(9\kappa^2)}\big)/8$, which proves (ii). Now, let $p = 4$. Then (4.8) asserts that π is the unique solution in $[0, +\infty[$ to the third degree equation $4\kappa\pi^3 + \pi - \xi = 0$, namely $\pi = \big(\sqrt{\alpha^2 + \beta^3} - \alpha\big)^{1/3} - \big(\sqrt{\alpha^2 + \beta^3} + \alpha\big)^{1/3}$, where $\alpha = -\xi/(8\kappa)$ and $\beta = 1/(12\kappa)$. Since this expression is an odd function of ξ, we obtain (iv). Finally, we deduce (i) from (iv) by observing that, since $\big(3|\cdot|^{4/3}/4\big)^* = |\cdot|^4/4$, Lemma 3.3(iii) with $\gamma = 4\kappa/3$ yields $\pi = \mathrm{prox}_{\gamma(3|\cdot|^{4/3}/4)} \xi = \xi - \gamma\, \mathrm{prox}_{(4\gamma)^{-1}|\cdot|^4}(\xi/\gamma)$, hence the result after simple algebra. ◻

Example 4.6 (Huber distribution) Let $\omega \in \,]0, +\infty[$, $\tau \in \,]0, +\infty[$, and set

$$\phi\colon \mathbb{R} \to \,]-\infty, +\infty] : \xi \mapsto \begin{cases} \tau\xi^2, & \text{if } |\xi| \leq \omega/\sqrt{2\tau};\\ \omega\sqrt{2\tau}|\xi| - \omega^2/2, & \text{otherwise.} \end{cases} \tag{4.9}$$

Then, for every $\xi \in \mathbb{R}$,

$$\mathrm{prox}_\phi \xi = \begin{cases} \dfrac{\xi}{2\tau + 1}, & \text{if } |\xi| \leq \omega(2\tau + 1)/\sqrt{2\tau};\\ \xi - \omega\sqrt{2\tau}\,\mathrm{sign}(\xi), & \text{if } |\xi| > \omega(2\tau + 1)/\sqrt{2\tau}. \end{cases} \tag{4.10}$$

Proof. Let $\xi \in \mathbb{R}$ and set $\pi = \mathrm{prox}_\phi \xi$. Since ϕ is even, we assume that $\xi \geq 0$ (see Remark 4.2(ii)). In addition, since ϕ is differentiable, it follows from (4.1) and (4.3) that π is the unique solution in $[0, \xi]$ to $\pi + \phi'(\pi) = \xi$. First, suppose that $\pi = \omega/\sqrt{2\tau}$. Then $\phi'(\pi) = \omega\sqrt{2\tau}$ and, therefore, $\xi = \pi + \phi'(\pi) = \omega(2\tau + 1)/\sqrt{2\tau}$. Now, suppose that $\xi \leq \omega(2\tau + 1)/\sqrt{2\tau}$. Then it follows from Lemma 4.1(i) that $\pi \leq \mathrm{prox}_\phi(\omega(2\tau+1)/\sqrt{2\tau}) = \omega/\sqrt{2\tau}$. In turn, (4.9) yields $\phi'(\pi) = 2\tau\pi$ and the identity $\xi = \pi + \phi'(\pi)$ yields $\pi = \xi/(2\tau + 1)$. Finally, if $\xi > \omega(2\tau + 1)/\sqrt{2\tau}$, then Lemma 4.1(i) yields $\pi \geq \mathrm{prox}_\phi(\omega(2\tau + 1)/\sqrt{2\tau}) = \omega/\sqrt{2\tau}$ and, in turn, $\phi'(\pi) = \omega\sqrt{2\tau}$, which allows us to conclude that $\pi = \xi - \omega\sqrt{2\tau}$. ◻

Example 4.7 (maximum entropy distribution) This density is obtained by maximizing the entropy subject to the knowledge of the first, second, and p-th order absolute moments, where $2 \neq p \in \,]1, +\infty[$ [24]. Let $\omega \in \,]0, +\infty[$, $\tau \in [0, +\infty[$, $\kappa \in \,]0, +\infty[$, and set

$$\phi\colon \mathbb{R} \to \,]-\infty, +\infty] : \xi \mapsto \omega|\xi| + \tau|\xi|^2 + \kappa|\xi|^p. \tag{4.11}$$

Then, for every $\xi \in \mathbb{R}$,

$$\mathrm{prox}_\phi \xi = \mathrm{sign}(\xi)\, \mathrm{prox}_{\kappa|\cdot|^p/(2\tau+1)} \left(\frac{1}{2\tau + 1} \max\{|\xi| - \omega, 0\}\right) \tag{4.12}$$

where the expression of $\mathrm{prox}_{\kappa|\cdot|^p/(2\tau+1)}$ is supplied by Example 4.5.

Proof. The function ϕ is a quadratic perturbation of the function $\varphi = \omega|\cdot| + \kappa|\cdot|^p$. Applying Lemma 4.1(iii) with $\psi = \kappa|\cdot|^p$ and $\Omega = [-\omega, \omega]$, we get $(\forall \xi \in \mathbb{R})\ \mathrm{prox}_\varphi \xi = \mathrm{prox}_{\kappa|\cdot|^p}(\mathrm{soft}_{[-\omega,\omega]} \xi) = \mathrm{sign}(\xi)\,\mathrm{prox}_{\kappa|\cdot|^p}(\max\{|\xi| - \omega, 0\})$. Hence, the result follows from Lemma 3.2(i) where $\mathcal{X} = \mathbb{R}$, $\alpha = 2\tau$, and $\mathsf{u} = 0$. ◻

Example 4.8 (smoothed Laplace distribution) Let $\omega \in \left]0, +\infty\right[$ and set

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \omega|\xi| - \ln(1 + \omega|\xi|). \tag{4.13}$$

This potential function is sometimes used as a differentiable approximation to (4.4), e.g., [28]. We have, for every $\xi \in \mathbb{R}$,

$$\mathrm{prox}_\phi \xi = \mathrm{sign}(\xi)\frac{\omega|\xi| - \omega^2 - 1 + \sqrt{\left|\omega|\xi| - \omega^2 - 1\right|^2 + 4\omega|\xi|}}{2\omega}. \tag{4.14}$$

Proof. According to Remark 4.2(ii), since ϕ is even, we can focus on the case when $\xi \geq 0$. As ϕ achieves its infimum at 0, Lemma 4.1(ii) yields $\pi = \mathrm{prox}_\phi \xi \geq 0$. We deduce from (4.3) that π is the unique nonnegative solution to the equation

$$\omega\pi^2 + (\omega^2 + 1 - \omega\xi)\pi - \xi = 0, \tag{4.15}$$

which leads to (4.14). ◻

Example 4.9 (exponential distribution) Let $\omega \in \left]0, +\infty\right[$ and set

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} \omega\xi, & \text{if } \xi \geq 0; \\ +\infty, & \text{if } \xi < 0. \end{cases} \tag{4.16}$$

Then, for every $\xi \in \mathbb{R}$,

$$\mathrm{prox}_\phi \xi = \begin{cases} \xi - \omega & \text{if } \xi \geq \omega; \\ 0 & \text{if } \xi < \omega. \end{cases} \tag{4.17}$$

Proof. Set $\varphi = \iota_{[0,+\infty[}$. Then Definition 3.1 yields $\mathrm{prox}_\varphi = P_{[0,+\infty[}$. In turn, since ϕ is a linear perturbation of φ, the claim results from Lemma 3.2(i), where $\mathcal{X} = \mathbb{R}$, $\alpha = 0$, and $\mathsf{u} = \omega$. ◻

Example 4.10 (gamma distribution) Let $\omega \in \left]0, +\infty\right[$, $\kappa \in \left]0, +\infty\right[$, and set

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} -\kappa\ln(\xi) + \omega\xi, & \text{if } \xi > 0; \\ +\infty, & \text{if } \xi \leq 0. \end{cases} \tag{4.18}$$

Then, for every $\xi \in \mathbb{R}$,

$$\mathrm{prox}_\phi \xi = \frac{\xi - \omega + \sqrt{|\xi - \omega|^2 + 4\kappa}}{2}. \tag{4.19}$$

Proof. Set

$$\varphi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\kappa\ln(\xi), & \text{if } \xi > 0;\\ +\infty, & \text{if } \xi \leq 0. \end{cases} \tag{4.20}$$

We easily get from Remark 4.2(i)&(iii) that

$$(\forall \xi \in \mathbb{R}) \quad \operatorname{prox}_{\varphi} \xi = \frac{\xi + \sqrt{\xi^2 + 4\kappa}}{2}. \tag{4.21}$$

In turn, since ϕ is a linear perturbation of φ, the claim results from Lemma 3.2(i), where $\mathcal{X} = \mathbb{R}$, $\alpha = 0$, and $\mathsf{u} = \omega$. ☐

Example 4.11 (chi distribution) Let $\kappa \in \left]0,+\infty\right[$ and let

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\kappa\ln(\xi) + \xi^2/2, & \text{if } \xi > 0;\\ +\infty, & \text{if } \xi \leq 0. \end{cases} \tag{4.22}$$

Then, for every $\xi \in \mathbb{R}$,

$$\operatorname{prox}_{\phi} \xi = \frac{\xi + \sqrt{\xi^2 + 8\kappa}}{4}. \tag{4.23}$$

Proof. Since ϕ is a quadratic perturbation of the function φ defined in (4.20), the claim results from Lemma 3.2(i), where $\mathcal{X} = \mathbb{R}$, $\alpha = 1$, and $\mathsf{u} = 0$. ☐

Example 4.12 (uniform distribution) Let $\omega \in \left]0,+\infty\right[$ and set

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} 0, & \text{if } |\xi| \leq \omega;\\ +\infty, & \text{otherwise.} \end{cases} \tag{4.24}$$

Then it follows at once from Definition 3.1 that, for every $\xi \in \mathbb{R}$,

$$\operatorname{prox}_{\phi} \xi = P_{[-\omega,\omega]}\xi = \begin{cases} -\omega, & \text{if } \xi < -\omega;\\ \xi, & \text{if } |\xi| \leq \omega;\\ \omega, & \text{if } \xi > \omega. \end{cases} \tag{4.25}$$

Example 4.13 (triangular distribution) Let $\underline{\omega} \in \left]-\infty,0\right[$, let $\overline{\omega} \in \left]0,+\infty\right[$, and set

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\ln(\xi - \underline{\omega}) + \ln(-\underline{\omega}), & \text{if } \xi \in \left]\underline{\omega},0\right];\\ -\ln(\overline{\omega} - \xi) + \ln(\overline{\omega}), & \text{if } \xi \in \left]0,\overline{\omega}\right[;\\ +\infty, & \text{otherwise.} \end{cases} \tag{4.26}$$

Then, for every $\xi \in \mathbb{R}$,

$$\operatorname{prox}_{\phi} \xi = \begin{cases} \dfrac{\xi + \underline{\omega} + \sqrt{|\xi - \underline{\omega}|^2 + 4}}{2}, & \text{if } \xi < 1/\underline{\omega};\\ \dfrac{\xi + \overline{\omega} - \sqrt{|\xi - \overline{\omega}|^2 + 4}}{2}, & \text{if } \xi > 1/\overline{\omega};\\ 0 & \text{otherwise.} \end{cases} \tag{4.27}$$

Proof. Let $\xi \in \mathbb{R}$ and set $\pi = \operatorname{prox}_\phi \xi$. Let us first note that $\partial\phi(0) = [1/\underline{\omega}, 1/\overline{\omega}]$. Therefore, (3.3) yields

$$\pi = 0 \quad \Leftrightarrow \quad \xi \in [1/\underline{\omega}, 1/\overline{\omega}]\,. \tag{4.28}$$

Now consider the case when $\xi > 1/\overline{\omega}$. Since ϕ admits 0 as a minimizer, it follows from Lemma 4.1(ii) and (4.28) that $\pi \in \,]0,\xi]$. Hence, we derive from (4.3) that π is the only solution in $]0,\xi]$ to $\pi + 1/(\overline{\omega} - \pi) = \xi$, i.e., $\pi = (\xi + \overline{\omega} - \sqrt{|\xi - \overline{\omega}|^2 + 4})/2$. Likewise, if $\xi < 1/\underline{\omega}$, it follows from Lemma 4.1(ii), (4.28), and (4.3) that π is the only solution in $[\xi, 0[$ to $\pi - 1/(\pi - \underline{\omega}) = \xi$, which yields $\pi = (\xi + \underline{\omega} + \sqrt{|\xi - \underline{\omega}|^2 + 4})/2$. ☐

The next example is an extension of Example 4.11.

Example 4.14 (Weibull distribution) Let $\omega \in \,]0,+\infty[$, $\kappa \in \,]0,+\infty[$, and $p \in \,]1,+\infty[$, and set

$$\phi\colon \mathbb{R} \to \,]-\infty,+\infty] : \xi \mapsto \begin{cases} -\kappa \ln(\xi) + \omega\xi^p, & \text{if } \xi > 0;\\ +\infty, & \text{if } \xi \leq 0. \end{cases} \tag{4.29}$$

Then, for every $\xi \in \mathbb{R}$, $\pi = \operatorname{prox}_\phi \xi$ is the unique strictly positive solution to

$$p\omega\pi^p + \pi^2 - \xi\pi = \kappa. \tag{4.30}$$

Proof. Since ϕ is differentiable on $]0,+\infty[$, it follows from Remark 4.2(i)&(iii) that π is the unique solution in $]0,+\infty[$ to $\pi + \phi'(\pi) = \xi$ or, equivalently, to (4.30). ☐

A similar proof can be used in the following two examples.

Example 4.15 (generalized inverse Gaussian distribution) Let $\omega \in \,]0,+\infty[$, $\kappa \in [0,+\infty[$, and $\rho \in \,]0,+\infty[$, and set

$$\phi\colon \mathbb{R} \to \,]-\infty,+\infty] : \xi \mapsto \begin{cases} -\kappa \ln(\xi) + \omega\xi + \rho/\xi, & \text{if } \xi > 0;\\ +\infty, & \text{if } \xi \leq 0. \end{cases} \tag{4.31}$$

Then, for every $\xi \in \mathbb{R}$, $\pi = \operatorname{prox}_\phi \xi$ is the unique strictly positive solution to

$$\pi^3 + (\omega - \xi)\pi^2 - \kappa\pi = \rho. \tag{4.32}$$

Example 4.16 (Pearson type I) Let $\underline{\kappa}$ and $\overline{\kappa}$ be in $]0,+\infty[$, let $\underline{\omega}$ and $\overline{\omega}$ be nonnegative reals such that $\underline{\omega} < \overline{\omega}$, and set

$$\phi\colon \mathbb{R} \to \,]-\infty,+\infty] : \xi \mapsto \begin{cases} -\underline{\kappa} \ln(\xi - \underline{\omega}) - \overline{\kappa} \ln(\overline{\omega} - \xi), & \text{if } \xi \in \,]\underline{\omega}, \overline{\omega}[\,;\\ +\infty, & \text{otherwise.} \end{cases} \tag{4.33}$$

Then, for every $\xi \in \mathbb{R}$, $\pi = \operatorname{prox}_\phi \xi$ is the unique solution in $]\underline{\omega}, \overline{\omega}[$ to

$$\pi^3 - (\underline{\omega} + \overline{\omega} + \xi)\pi^2 + \big(\underline{\omega}\overline{\omega} - \underline{\kappa} - \overline{\kappa} + (\underline{\omega} + \overline{\omega})\xi\big)\pi = \underline{\omega}\overline{\omega}\xi - \underline{\omega}\overline{\kappa} - \overline{\omega}\underline{\kappa}. \tag{4.34}$$

Remark 4.17

(i) The chi-square distribution with $n > 2$ degrees of freedom is a special case of the gamma distribution (Example 4.10) with $(\omega, \kappa) = (1/2, n/2 - 1)$.

(ii) The normalized Rayleigh distribution is a special case of the chi distribution (Example 4.11) with $\kappa = 1$.

(iii) The beta distribution and the Wigner distribution are special cases of the Pearson type I distribution (Example 4.16) with $(\underline{\omega}, \overline{\omega}) = (0, 1)$, and $-\underline{\omega} = \overline{\omega}$ and $\underline{\kappa} = \overline{\kappa} = 1/2$, respectively.

(iv) The proximity operator associated with translated and/or scaled versions of the above densities can be obtained via Lemma 3.2(ii)&(iii).

(v) For log-concave densities for which the proximity operator of the potential function is difficult to express in closed form (e.g., Kumaraswamy or logarithmic distributions), one can turn to simple procedures to solve (3.3) or (4.3) numerically.

5 Existence of solutions

We address the issue of the existence of solutions to Problem 2.1. Recall that a function $\varphi \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ is said to be coercive if $\lim_{\|x\| \to +\infty} \varphi(x) = +\infty$.

Proposition 5.1 *Suppose that one of the following holds.*

(i) *The function* $(\xi_k)_{k \in \mathbb{K}} \mapsto \sum_{k \in \mathbb{K}} \phi_k(\xi_k) + \Psi(F^*(\xi_k)_{k \in \mathbb{K}})$ *is coercive.*

(ii) $\inf_{k \in \mathbb{K}} \inf \phi_k(\mathbb{R}) > -\infty$, Ψ *is coercive, and* $(e_k)_{k \in \mathbb{K}}$ *is a Riesz basis.*

(iii) $\inf \Psi(\mathcal{H}) > -\infty$ *and one of the following properties is satisfied.*

   (a) *The function* $(\xi_k)_{k \in \mathbb{K}} \mapsto \sum_{k \in \mathbb{K}} \phi_k(\xi_k)$ *is coercive.*

   (b) *There exists* $\omega \in \left]0, +\infty\right[$ *and* $p \in [1, 2]$ *such that* $(\forall k \in \mathbb{K})\ \phi_k \geq \omega |\cdot|^p$.

   (c) $\mathbb{K}$ *is finite and the functions* $(\phi_k)_{k \in \mathbb{K}}$ *are coercive.*

Then Problem 2.1 admits a solution.

Proof. We denote by $\mathsf{x} = (\xi_k)_{k \in \mathbb{K}}$ a generic element in $\ell^2(\mathbb{K})$ and by $\|\mathsf{x}\| = \sqrt{\sum_{k \in \mathbb{K}} |\xi_k|^2}$ its norm. Set

$$f_1 \colon \mathsf{x} \mapsto \sum_{k \in \mathbb{K}} \phi_k(\xi_k) \quad \text{and} \quad f_2 = \Psi \circ F^*. \tag{5.1}$$

First, suppose that (i) holds. Then it follows from the assumptions on $(\phi_k)_{k \in \mathbb{K}}$ in Problem 2.1 and Proposition 3.6 that $f_1 \in \Gamma_0(\ell^2(\mathbb{K}))$. On the other hand, since Ψ is a finite function in

$\Gamma_0(\mathcal{H})$ and $F^*\colon \ell^2(\mathbb{K}) \to \mathcal{H}$ is linear and bounded, f_2 is a finite function in $\Gamma_0(\ell^2(\mathbb{K}))$. Altogether, $f_1 + f_2 \in \Gamma_0(\ell^2(\mathbb{K}))$ and the claim follows from [35, Theorem 2.5.1(ii)].

Next, suppose that (ii) holds. In view of (i), since f_1 is bounded below, it is enough to show that f_2 is coercive. Since $(e_k)_{k\in\mathbb{K}}$ is a Riesz basis, we have [26]

$$(\forall \mathsf{x} \in \ell^2(\mathbb{K})) \quad \|F^*\mathsf{x}\| \geq \sqrt{\mu}\|\mathsf{x}\|. \tag{5.2}$$

In turn, the coercivity of Ψ implies that $\lim_{\|\mathsf{x}\|\to+\infty} \|\Psi(F^*\mathsf{x})\| = +\infty$.

Now, suppose that (iii) holds. In case (iii)(a), since Ψ is bounded below, f_2 is likewise. In turn, the coercivity of f_1 implies that of $f_1 + f_2$, hence the result by (i). Now suppose that (iii)(b) is satisfied and let $\mathsf{x} \in \ell^2(\mathbb{K})$. Then

$$f_1(\mathsf{x}) = \sum_{k\in\mathbb{K}} \phi_k(\xi_k) \geq \omega \sum_{k\in\mathbb{K}} |\xi_k|^p \geq \omega\|\mathsf{x}\|^p. \tag{5.3}$$

Therefore f_1 is coercive and the claim follows from (iii)(a). Finally, suppose that (iii)(c) is satisfied. In view of (iii)(a), it is enough to show that f_1 is coercive. To this end, fix $\rho \in \left]0,+\infty\right[$ and recall that $\mathbb{K} = \{1,\ldots,K\}$. Since the functions $(\phi_k)_{k\in\mathbb{K}}$ are coercive, we can find $\delta \in \left]0,+\infty\right[$ such that

$$(\forall \xi \in \mathbb{R}) \quad |\xi| \geq \delta/\sqrt{K} \quad \Rightarrow \quad \min_{k\in\mathbb{K}} \phi_k(\xi) \geq \rho. \tag{5.4}$$

Now take $\mathsf{x} \in \ell^2(\mathbb{K})$ such that $\|\mathsf{x}\| \geq \delta$. Then $\max_{k\in\mathbb{K}} |\xi_k| \geq \delta/\sqrt{K}$ and therefore (5.4) yields $f_1(\mathsf{x}) \geq \rho$. ☐

6 Algorithm

We propose the following algorithm to solve Problem 2.1.

Algorithm 6.1 Fix $\mathsf{x}_0 \in \ell^2(\mathbb{K})$ and construct a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}} = ((\xi_{n,k})_{k\in\mathbb{K}})_{n\in\mathbb{N}}$ by setting, for every $n \in \mathbb{N}$,

$$(\forall k \in \mathbb{K}) \quad \xi_{n+1,k} = \xi_{n,k} + \lambda_n\Big(\operatorname{prox}_{\gamma_n\phi_k}\big(\xi_{n,k} - \gamma_n(\eta_{n,k} + \beta_{n,k})\big) + \alpha_{n,k} - \xi_{n,k}\Big), \tag{6.1}$$

where $\lambda_n \in \left]0,1\right]$, $\gamma_n \in \left]0,+\infty\right[$, $\{\alpha_{n,k}\}_{k\in\mathbb{K}} \subset \mathbb{R}$, $(\eta_{n,k})_{k\in\mathbb{K}} = F(\nabla\Psi(F^*\mathsf{x}_n))$, and $(\beta_{n,k})_{k\in\mathbb{K}} = Fb_n$, where $b_n \in \mathcal{H}$.

The chief advantage of this algorithm is to be fully split in the sense that the functions $(\phi_k)_{k\in\mathbb{K}}$ and Ψ appearing in (2.2) are used separately. First, the current iterate x_n is transformed into a point in $F^*\mathsf{x}_n$ in $\mathcal{H}$, and the gradient of Ψ is evaluated at this point to within some tolerance b_n. Next, one forms the sequences $(\beta_{n,k})_{k\in\mathbb{K}} = Fb_n$ and $(\eta_{n,k})_{k\in\mathbb{K}} = F(\nabla\Psi(F^*\mathsf{x}_n))$. Then one chooses $\gamma_n > 0$, and, for every $k \in \mathbb{K}$, applies the operator $\operatorname{prox}_{\gamma_n\phi_k}$ to $\xi_{n,k} - \gamma_n(\eta_{n,k} + \beta_{n,k})$. An error $\alpha_{n,k}$ is tolerated in this computation. Finally, the kth component $\xi_{n+1,k}$ of x_{n+1} is obtained by applying a relaxation of parameter λ_n to this inexact proximal step. Let us note that the computation of the proximal steps can be performed in parallel.

To study the asymptotic behavior of the sequences generated by Algorithm 6.1, we require the following set of assumptions.

Assumption 6.2 In addition to the standing assumptions of Problem 2.1, the following hold.

(i) Problem 2.1 admits a solution.

(ii) $\inf_{n\in\mathbb{N}} \lambda_n > 0$.

(iii) $\inf_{n\in\mathbb{N}} \gamma_n > 0$ and $\sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$, where β is a Lipschitz constant of $F \circ \nabla\Psi \circ F^*$.

(iv) $\sum_{n\in\mathbb{N}} \sqrt{\sum_{k\in\mathbb{K}} |\alpha_{n,k}|^2} < +\infty$ and $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$.

Remark 6.3 As regards Assumption 6.2(i), sufficient conditions can be found in Proposition 5.1. Let us now turn to the parameter β in Assumption 6.2(iii), which determines the range of the step sizes $(\gamma_n)_{n\in\mathbb{N}}$. It follows from the assumptions of Problem 2.1 and (1.1) that, for every x and y in $\ell^2(\mathbb{K})$,

$$\begin{aligned}\|F(\nabla\Psi(F^*\mathsf{x})) - F(\nabla\Psi(F^*\mathsf{y}))\| &\leq \|F\|\,\|\nabla\Psi(F^*\mathsf{x}) - \nabla\Psi(F^*\mathsf{y})\| \\ &\leq \tau\|F\|\,\|F^*\mathsf{x} - F^*\mathsf{y}\| \\ &\leq \tau\|F\|^2\,\|\mathsf{x} - \mathsf{y}\| \\ &\leq \tau\nu\|\mathsf{x} - \mathsf{y}\|. \end{aligned} \tag{6.2}$$

Thus, the value $\beta = \tau\nu$ can be used in general. In some cases, however, a sharper bound can be obtained, which results in a wider range for the step sizes $(\gamma_n)_{n\in\mathbb{N}}$. For example, in the problem considered in Section 2.2, if the norm of $R = \sum_{i=1}^{q} \alpha_i F T_i^* T_i F^*$ can be evaluated, it follows from (2.4) and the nonexpansivity of the operators $(\mathrm{Id} - P_{S_i})_{1\leq i\leq m}$ that one can take

$$\beta = \|R\| + \nu \sum_{i=1}^{m} \beta_i. \tag{6.3}$$

Theorem 6.4 *Let $(\mathsf{x}_n)_{n\in\mathbb{N}}$ be an arbitrary sequence generated by Algorithm 6.1 under Assumption 6.2. Then $(\mathsf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a solution to Problem 2.1.*

Proof. Set $\mathcal{X} = \ell^2(\mathbb{K})$ and let $(\mathsf{o}_k)_{k\in\mathbb{K}}$ denote the canonical orthonormal basis of $\ell^2(\mathbb{K})$. Then (2.2) can be written as

$$\underset{\mathsf{x}\in\mathcal{X}}{\text{minimize}} \;\sum_{k\in\mathbb{K}} \phi_k(\langle \mathsf{x} \mid \mathsf{o}_k\rangle) + \Psi(F^*\mathsf{x}). \tag{6.4}$$

Now set $f_1 = \sum_{k\in\mathbb{K}} \phi_k(\langle \cdot \mid \mathsf{o}_k\rangle)$ and $f_2 = \Psi \circ F^*$. Then, in the light of the assumptions on $(\phi_k)_{k\in\mathbb{K}}$ in Problem 2.1, Proposition 3.6 yields $f_1 \in \Gamma_0(\mathcal{X})$. On the other hand, since Ψ is a finite function in $\Gamma_0(\mathcal{H})$ and $F^*\colon \mathcal{X} \to \mathcal{H}$ is linear and bounded, we have $f_2 \in \Gamma_0(\mathcal{X})$. In addition, $\nabla f_2 = F \circ \nabla\Psi \circ F^*$ is β-Lipschitz continuous (see Assumption 6.2(iii)). Altogether, (6.4) conforms to the format of Problem 3.4. Furthermore, it follows from Proposition 3.6 that we can rewrite (6.1) as

$$\begin{aligned}\mathsf{x}_{n+1} &= \mathsf{x}_n + \lambda_n\bigg(\sum_{k\in\mathbb{K}} \big(\mathrm{prox}_{\gamma_n\phi_k} \langle \mathsf{x}_n - \gamma_n F(\nabla\Psi(F^*\mathsf{x}_n) + b_n) \mid \mathsf{o}_k\rangle + \alpha_{n,k}\big)\mathsf{o}_k - \mathsf{x}_n\bigg) \\ &= \mathsf{x}_n + \lambda_n\Big(\mathrm{prox}_{\gamma_n f_1} \big(\mathsf{x}_n - \gamma_n(\nabla f_2(\mathsf{x}_n) + \mathsf{b}_n)\big) + \mathsf{a}_n - \mathsf{x}_n\Big), \end{aligned} \tag{6.5}$$

where $\mathsf{a}_n = (\alpha_{n,k})_{k\in\mathbb{K}}$ and $\mathsf{b}_n = Fb_n$. Since Assumption 6.2(iv) and (1.1) imply that $\sum_{n\in\mathbb{N}} \|\mathsf{a}_n\| < +\infty$ and $\sum_{n\in\mathbb{N}} \|\mathsf{b}_n\| \leq \sqrt{\nu}\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$, the claim therefore follows from Theorem 3.5. ☐

Let $(\mathsf{x}_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 6.1 under Assumption 6.2 and set $(\forall n \in \mathbb{N})$ $x_n = F^*\mathsf{x}_n$. On the one hand, Theorem 6.4 asserts that $(\mathsf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a solution x to Problem 2.1. On the other hand, since F^* is linear and bounded, it is weakly continuous and, therefore, $(x_n)_{n\in\mathbb{N}}$ converges weakly to $F^*\mathsf{x}$. However, it is not possible to express (6.1) as an iteration in terms of the sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ in general. The following corollary addresses the case when F is surjective, which does lead to an algorithm in $\mathcal{H}$.

Corollary 6.5 *Suppose that $(e_k)_{k\in\mathbb{K}}$ is a Riesz basis with companion biorthogonal basis $(\breve{e}_k)_{k\in\mathbb{K}}$. Fix $x_0 \in \mathcal{H}$ and, for every $n \in \mathbb{N}$, set*

$$x_{n+1} = x_n + \lambda_n\left(\sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\gamma_n\phi_k}\big(\langle x_n \mid \breve{e}_k\rangle - \gamma_n\langle\nabla\Psi(x_n) + b_n \mid e_k\rangle\big) + \alpha_{n,k}\big)\, e_k - x_n\right), \tag{6.6}$$

where $\lambda_n \in \,]0,1]$, $\gamma_n \in \,]0,+\infty[$, $\{\alpha_{n,k}\}_{k\in\mathbb{K}} \subset \mathbb{R}$, and $b_n \in \mathcal{H}$. Suppose that Assumption 6.2 is in force. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point $x \in \mathcal{H}$ and $(\langle x \mid \breve{e}_k\rangle)_{k\in\mathbb{K}}$ is a solution to Problem 2.1.

Proof. Set $(\forall n \in \mathbb{N})(\forall k \in \mathbb{K})$ $\xi_{n,k} = \langle x_n \mid \breve{e}_k\rangle$, $\eta_{n,k} = \langle\nabla\Psi(x_n) \mid e_k\rangle$, and $\beta_{n,k} = \langle b_n \mid e_k\rangle$. Then, for every $n \in \mathbb{N}$, it follows from (1.4) that $x_n = F^*(\xi_{n,k})_{k\in\mathbb{K}}$ and, in turn, that

$$(\eta_{n,k})_{k\in\mathbb{K}} = F(\nabla\Psi(F^*\mathsf{x}_n)), \quad \text{where} \quad \mathsf{x}_n = (\xi_{n,k})_{k\in\mathbb{K}}. \tag{6.7}$$

Furthermore, for every $n \in \mathbb{N}$, it follows from (6.6) and the biorthogonality of $(e_k)_{k\in\mathbb{K}}$ and $(\breve{e}_k)_{k\in\mathbb{K}}$ that

$$\begin{aligned}(\forall k \in \mathbb{K}) \quad \xi_{n+1,k} &= \langle x_{n+1} \mid \breve{e}_k\rangle \\ &= \langle x_n \mid \breve{e}_k\rangle + \lambda_n\Big(\operatorname{prox}_{\gamma_n\phi_k}\big(\langle x_n \mid \breve{e}_k\rangle - \gamma_n\langle\nabla\Psi(x_n) + b_n \mid e_k\rangle\big) + \alpha_{n,k} \\ &\qquad\qquad - \langle x_n \mid \breve{e}_k\rangle\Big) \\ &= \xi_{n,k} + \lambda_n\Big(\operatorname{prox}_{\gamma_n\phi_k}\big(\xi_{n,k} - \gamma_n(\eta_{n,k} + \beta_{n,k})\big) + \alpha_{n,k} - \xi_{n,k}\Big). \end{aligned} \tag{6.8}$$

Since Theorem 6.4 states that $(\mathsf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a solution x to Problem 2.1, $(x_n)_{n\in\mathbb{N}} = (F^*\mathsf{x}_n)_{n\in\mathbb{N}}$ converges weakly to $x = F^*\mathsf{x}$. Consequently, (1.4) asserts that we can write $\mathsf{x} = (\langle x \mid \breve{e}_k\rangle)_{k\in\mathbb{K}}$. ☐

Remark 6.6 Suppose that $\mathbb{K} = \mathbb{N}$ and that $(e_k)_{k\in\mathbb{K}}$ is an orthonormal basis of $\mathcal{H}$. Then (6.6) reduces to

$$x_{n+1} = x_n + \lambda_n\left(\sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\gamma_n\phi_k}\big(\langle x_n - \gamma_n(\nabla\Psi(x_n) + b_n) \mid e_k\rangle\big) + \alpha_{n,k}\big)\, e_k - x_n\right). \tag{6.9}$$

In this particular setting, some results related to Corollary 6.5 are the following.

(i) Suppose that $\Psi\colon x \mapsto \|Tx - z\|^2/2$, where T is a nonzero bounded linear operator from $\mathcal{H}$ to a real Hilbert space $\mathcal{G}$ and $z \in \mathcal{G}$. Suppose that, in addition, $(\forall k \in \mathbb{K})\ \phi_k \geq \phi_k(0) = 0$. Then the convergence of (6.9) is discussed in [15, Corollary 5.16].

(ii) Suppose that $(\Omega_k)_{k\in\mathbb{K}}$ are closed intervals of $\mathbb{R}$ such that $0 \in \operatorname{int} \bigcap_{k\in\mathbb{K}} \Omega_k$ and that

$$(\forall k \in \mathbb{K}) \quad \phi_k = \psi_k + \sigma_{\Omega_k}, \tag{6.10}$$

where $\psi_k \in \Gamma_0(\mathbb{R})$ is differentiable at 0 and $\psi_k \geq \psi_k(0) = 0$. Then (6.9) is the thresholding algorithm proposed and analyzed in [14], namely

$$x_{n+1} = x_n + \lambda_n \Bigg(\sum_{k\in\mathbb{K}} \Big(\operatorname{prox}_{\gamma_n \psi_k} \big(\operatorname{soft}_{\gamma_n \Omega_k} \langle x_n - \gamma_n(\nabla\Psi(x_n) + b_n) \mid e_k \rangle \big) + \alpha_{n,k} \Big) e_k - x_n \Bigg), \tag{6.11}$$

where $\operatorname{soft}_{\gamma_n \Omega_k}$ is defined in (4.2).

(iii) Suppose that the assumptions of both (i) and (ii) hold and that, in addition, we set $\lambda_n \equiv 1$, $\|T\| < 1$, $\gamma_n \equiv 1$, $\alpha_{n,k} \equiv 0$, $b_n \equiv 0$, and $(\forall k \in \mathbb{K})\ \psi_k = 0$ and $\Omega_k = [-\omega_k, \omega_k]$. Then (6.11) becomes

$$x_{n+1} = \sum_{k\in\mathbb{K}} \big(\operatorname{soft}_{\Omega_k} \langle x_n + T^*(z - Tx_n) \mid e_k \rangle \big) e_k. \tag{6.12}$$

Algorithm 6.1 can be regarded as a descendant of this original method, which is investigated in [17] and [18].

7 Numerical results

The proposed framework is applicable to a wide array of variational formulations for inverse problems over frames. We provide a couple of examples to illustrate its applicability in wavelet-based image restoration in the Euclidean space $\mathcal{H} = \mathbb{R}^{512\times 512}$. The choice of the potential functions $(\phi_k)_{k\in\mathbb{K}}$ in Problem 2.1 is guided by the observation that regular images typically possess sparse wavelet representations and that the resulting wavelet coefficients often have even probability density functions [26]. Among the candidate potential functions investigated in Section 4, those of Example 4.7 appear to be the most appropriate for modeling wavelet coefficients on two counts. First, they provide flexible models of even potentials. Second, as shown in Lemma 4.1(iii), their proximity operators are thresholders and they therefore promote sparsity. More precisely, we employ potential functions of the form $\phi_k = \omega_k|\cdot| + \tau_k|\cdot|^2 + \kappa_k|\cdot|^{p_k}$, where $p_k \in \{4/3, 3/2, 3, 4\}$ and $\{\omega_k, \tau_k, \kappa_k\} \subset \left]0, +\infty\right[$. Note that $\operatorname{prox}_{\phi_k}$ can be obtained explicitly via (4.12) and Examples 4.5(i)-(iv). In addition, it follows from Proposition 5.1(iii)(b) that, with such potential functions, Problem 2.1 does admit a solution. The values of the parameters ω_k, τ_k, and κ_k are chosen for each wavelet subband via a maximum likelihood approach. The first example uses a biorthogonal wavelet frame and the second one uses an M-band dual-tree wavelet frame. Let us emphasize that such decompositions cannot be dealt with using the methods developed in [14], which are limited to orthonormal basis representations.

7.1 Example 1

We provide a multiview restoration example in a biorthogonal wavelet basis. The original image $\overline{x}$ is the standard test image displayed in Fig. 1 (top left). Two observations (see Fig. 1 top right and bottom left) conforming to the model (2.3) are available. In our experiment, $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{H}$ and v_1 and v_2 are realizations of two independent zero-mean Gaussian white noise processes. Moreover, the operator T_1 models a motion blur in the diagonal direction and satisfies $\|T_1\| = 1$, whereas $T_2 = \mathrm{Id}\,/2$. The blurred image-to-noise ratio is higher for the first observation (22.79 dB versus 15.18 dB) and so is the relative error (18.53 dB versus 5.891 dB) (the decibel value of the relative error between an image z and $\overline{x}$ is $20\log_{10}(\|\overline{x}\|/\|z-\overline{x}\|)$). The function Ψ in Problem 2.1 is given by (2.4), where $\alpha_1 = 4.00 \times 10^{-2}$ and $\alpha_2 = 6.94 \times 10^{-3}$. In addition, we set $m = 1$, $\beta_1 = 10^{-2}$, and $S_1 = [0, 255]^{512\times512}$ to enforce the known range of the pixel values. A discrete biorthogonal spline 9-7 decomposition [2] is used over 3 resolution levels. Algorithm 6.1 with $\lambda_n \equiv 1$ is used to solve Problem 2.1. By numerically evaluating $\|R\|$ in (6.3), we obtain $\beta = 0.230$ and the step sizes are chosen to be $\gamma_n \equiv 1.99/\beta = 8.66$. The resulting restored image, shown in Fig. 1 (bottom right), yields a relative error of 23.84 dB.

7.2 Example 2

The original SPOT5 satellite image $\overline{x}$ is shown in Fig. 2 (top) and the degraded image z in $\mathcal{G} = \mathcal{H}$ is shown in Fig. 2 (center). The degradation model is given by (2.7), where T is a 7×7 uniform blur with $\|T\| = 1$, and where v is a realization of a zero-mean Gaussian white noise process. The blurred image-to-noise ratio is 28.08 dB and the relative error is 12.49 dB.

In this example, we perform a restoration in a discrete two-dimensional version of an M-band dual-tree wavelet frame [7]. This decomposition has a redundancy factor of 2 (i.e., with the notation of Section 2.3, $K/N = 2$). In our experiments, decompositions over 2 resolution levels are performed with $M = 4$ using the filter bank proposed in [1]. The function Ψ in Problem 2.1 is given by (2.13), where f_V is the probability density function of the Gaussian noise. A solution is obtained via Algorithm 6.1 with $\lambda_n \equiv 1$. For the representation under consideration, we derive from (6.3) that $\beta = 2$ and we set $\gamma_n \equiv 0.995$. The restored image, shown in Fig. 2 (bottom), yields a relative error of 15.68 dB, i.e., a significant improvement of over 3 dB in terms of signal-to-noise ratio. A more precise inspection of the magnified areas displayed in Fig. 3 shows that the proposed method makes it possible to recover sharp edges while removing noise in uniform areas.

References

[1] O. Alkin and H. Caglar, Design of efficient M-band coders with linear-phase and perfect-reconstruction properties, *IEEE Trans. Signal Process.,* vol. 43, pp. 1579–1590, 1995.

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image coding using wavelet transform, *IEEE Trans. Image Process.,* vol. 1, pp. 205–220, 1992.

[3] J. M. Bioucas-Dias, Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors, *IEEE Trans. Image Process.,* vol. 15, pp. 937–951, 2006.

[4] C. Bouman and K. Sauer, A generalized Gaussian image model for edge-preserving MAP estimation, *IEEE Trans. Image Process.,* vol. 2, pp. 296–310, 1993.

[5] C. L. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction, *Inverse Problems,* vol. 20, pp. 103–120, 2004.

[6] E. J. Candès and D. L. Donoho, Recovering edges in ill-posed inverse problems: Optimality of curvelet frames, *Ann. Statist.,* vol. 30, pp. 784–842, 2002.

[7] C. Chaux, L. Duval, and J.-C. Pesquet, Image analysis using a dual-tree M-band wavelet transform, *IEEE Trans. Image Process.,* vol. 15, pp. 2397–2412, 2006.

[8] A. Cohen, *Numerical Analysis of Wavelet Methods.* Elsevier, New York, 2003.

[9] A. Cohen, I. Daubechies, and J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.,* vol. 45, pp. 485–560, 1992.

[10] P. L. Combettes, The foundations of set theoretic estimation, *Proc. IEEE,* vol. 81, pp. 182–208, 1993.

[11] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Trans. Signal Process.,* vol. 42, pp. 2955–2966, 1994.

[12] P. L. Combettes, A block-iterative surrogate constraint splitting method for quadratic signal recovery, *IEEE Trans. Signal Process.,* vol. 51, pp. 1771–1782, 2003.

[13] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization,* vol. 53, pp. 475–504, 2004.

[14] P. L. Combettes and J.-C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, submitted. This paper can be downloaded from the *Optimization on-line* site: `http://www.optimization-online.org/DB_HTML/2006/09/1470.html`.

[15] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.,* vol. 4, pp. 1168–1200, 2005.

[16] I. Daubechies, *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992.

[17] 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.

[18] C. de Mol and M. Defrise, A note on wavelet-based inversion algorithms, *Contemp. Math.*, vol. 313, pp. 85–96, 2002.

[19] L. Devroye, *Non-Uniform Random Variate Generation.* Springer-Verlag, New York, 1986.

[20] M. N. Do and M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, *IEEE Trans. Image Process.,* vol. 14, pp. 2091–2106, 2005.

[21] M. Elad and A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, *IEEE Trans. Image Process.,* vol. 6, pp. 1646–1658, 1997.

[22] O. D. Escoda, L. Granai, and P. Vandergheynst, On the use of a priori information for sparse signal approximations, *IEEE Trans. Signal Process.,* vol. 54, pp. 3468–3482, 2006.

[23] D. Han and D. R. Larson, *Frames, Bases, and Group Representations.* Mem. Amer. Math. Soc., vol. 147, 2000.

[24] J. N. Kapur and H. K. Kesevan, *Entropy Optimization Principles with Applications.* Academic Press, Boston, 1992.

[25] E. Le Pennec and S. G. Mallat, Sparse geometric image representations with bandelets, *IEEE Trans. Image Process.,* vol. 14, pp. 423–438, 2005.

[26] S. G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, New York, 1999.

[27] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France,* vol. 93, pp. 273-299, 1965.

[28] M. Nikolova and M. K. Ng, Analysis of half-quadratic minimization methods for signal and image recovery, *SIAM J. Sci. Comput.,* vol. 27, pp. 937–966, 2005.

[29] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, The dual-tree complex wavelet transform, *IEEE Signal Process. Mag.*, vol. 22, pp. 123–151, 2005.

[30] H. Stark, ed., *Image Recovery: Theory and Application.* Academic Press, San Diego, CA, 1987.

[31] A. M. Thompson and J. Kay, On some Bayesian choices of regularization parameter in image restoration, *Inverse Problems,* vol. 9, pp. 749–761, 1993.

[32] J. A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inform. Theory,* vol. 52, pp. 1030–1051, 2006.

[33] R. Tolimieri and M. An, *Time-Frequency Representations.* Birkhäuser, Boston, MA, 1998.

[34] H. L. Van Trees, *Detection, Estimation, and Modulation Theory – Part I.* Wiley, New York, 1968.

[35] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World Scientific, River Edge, NJ, 2002.

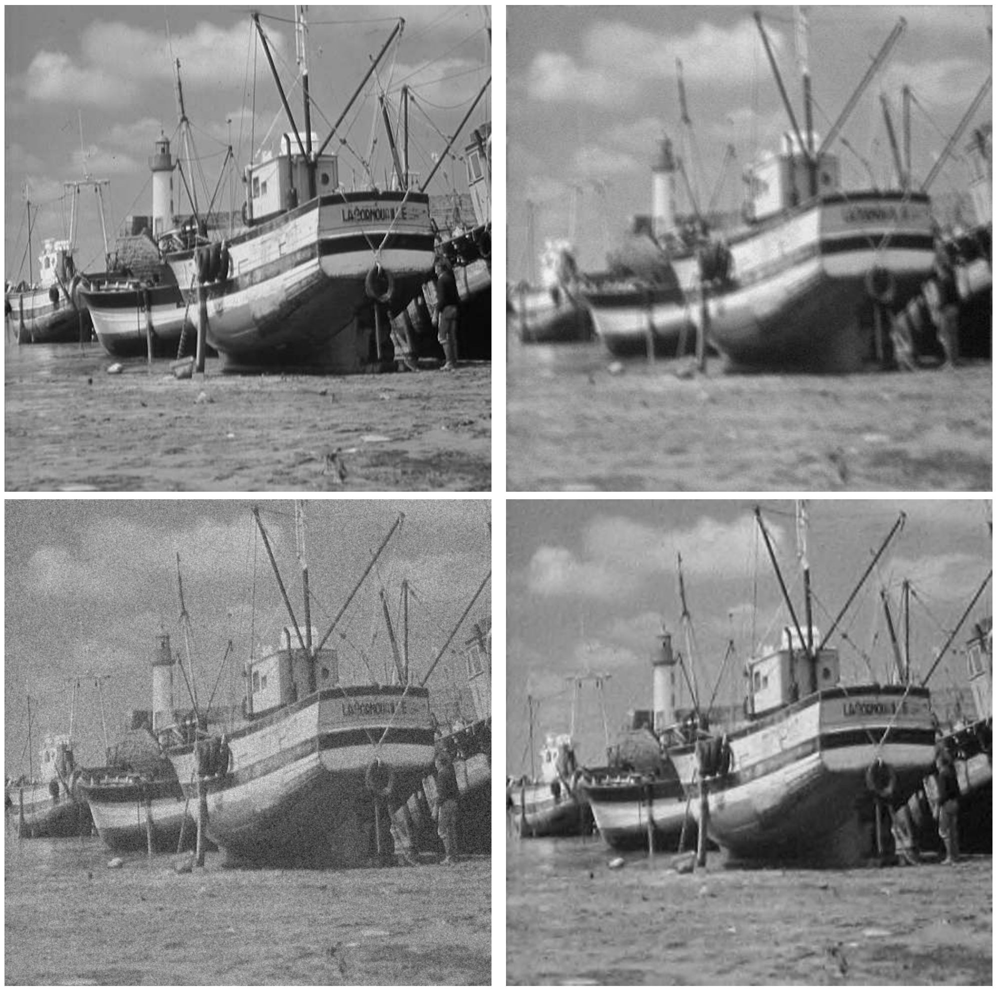

Figure 1: Example 1 – Original image (top left), first observation (top right), second observation (bottom left), and restored image (bottom right).

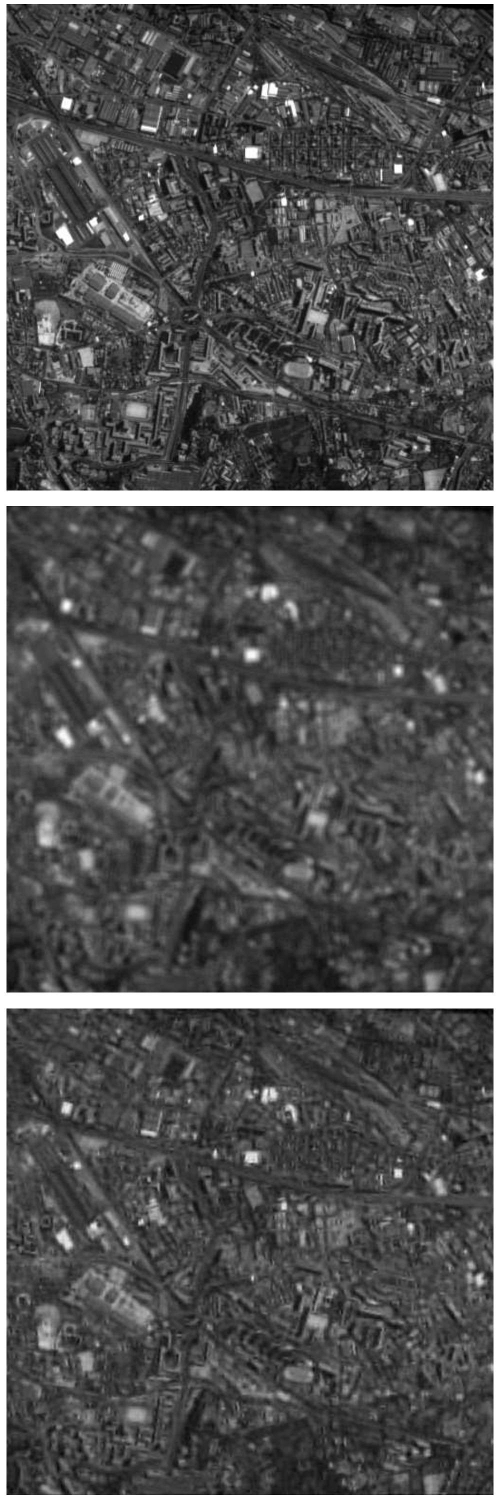

Figure 2: Example 2 – Original image (top); degraded image (center); image restored in a dual-tree wavelet frame (bottom).

Figure 3: Example 2 – Zoom on a 100×100 portion of the SPOT5 satellite image. Original image (top); degraded image (center); image restored in a dual-tree wavelet frame (bottom).