Proximal Comixture Minimization Models for Image Recovery and Data Analysis*

Patrick L. Combettes[1] and Diego J. Cornejo[2]

[1]North Carolina State University
Department of Mathematics
Raleigh, NC 27695, USA
plc@math.ncsu.edu

[2]Universidad del Pacífico
Lima, Perú
dj.cornejok@up.edu.pe

Abstract. In minimization models for image recovery and data analysis problems, loss functions and linear operators are typically aggregated as an average of composite terms. Each term in the aggregate models a desired property of the ideal solution arising from the *a priori* knowledge and the observed data. We propose an alternative minimization model based on proximal comixtures, an operation which combines functions and linear operators in such a way that the proximity operator of the resulting function is computable explicitly in terms of the individual proximity and linear operators. The mathematical properties of this operation are analyzed and comparisons between proximal comixtures and standard composite averages are made. Numerical illustrations of the benefits of minimization models based on proximal comixtures are provided in the context of image recovery and machine learning applications.

Keywords. Convex minimization, image recovery, Moreau envelope, proximal average, proximal cocomposition, proximal comixture.

---

*Contact author: P. L. Combettes. Email: plc@math.ncsu.edu. Phone: +1 919 515 2671.

This work was supported by the National Science Foundation under grant CCF-2211123.

A preliminary version of this work appears in the conference paper [17] with a simpler comixture optimization model and without any theoretical analysis.

§1. Introduction

The objective of image recovery and, more broadly, of various tasks in the areas of inverse problems and data analysis is to determine the value of a mathematical object (an image, a signal, a set of parameters, a distribution, a covariance matrix, a spectrum, etc.) conveying information of interest using experimental measurements and some *a priori* knowledge (data formation model, properties of the target solution, etc.). Over the past 60 years, convex optimization has established itself as one of the most reliable and efficient framework to formulate, analyze, and solve such problems [1, 11, 20, 22, 23, 25, 28, 32, 39, 45]. One typically aims at minimizing an aggregation of convex loss functions that model individually the desired properties of the ideal solution arising from the *a priori* knowledge and the observed data. The assumptions underlying the minimization models to be discussed are the following (see Section 2 for notation).

Assumption 1.1. $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle_{\mathcal{H}}$, associated norm $\|\cdot\|_{\mathcal{H}}$, and quadratic kernel $\mathscr{Q}_{\mathcal{H}} = \|\cdot\|_{\mathcal{H}}^2/2$, $f \in \Gamma_0(\mathcal{H})$, and $h\colon \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with a β^{-1}-Lipschitzian gradient for some $\beta \in \left]0,+\infty\right[$. Further, $0 < p \in \mathbb{N}$ and, for every $k \in \{1,\ldots,p\}$, $\mathcal{G}_k$ is a real Hilbert space, $g_k \in \Gamma_0(\mathcal{G}_k)$, and $0 \neq L_k\colon \mathcal{H} \to \mathcal{G}_k$ is a bounded linear operator. Finally, $(\alpha_k)_{1\leqslant k\leqslant p}$ are weights in $\left]0,+\infty\right[$ which satisfy $\sum_{k=1}^{p} \alpha_k \|L_k\|^2 \leqslant 1$.

The most prevalent minimization framework in data analysis consists in minimizing an objective function which aggregates the loss functions $(g_k)_{1\leqslant k\leqslant p}$ and the linear operators $(L_k)_{1\leqslant k\leqslant p}$ by means of a standard composite averaging operation as follows.

Problem 1.2. Under Assumption 1.1, the task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \big(\mathrm{cav}(g_k, L_k)_{1\leqslant k\leqslant p}\big)(x) + h(x), \tag{1.1}$$

where

$$\mathrm{cav}(g_k, L_k)_{1\leqslant k\leqslant p} = \sum_{k=1}^{p} \alpha_k g_k \circ L_k \tag{1.2}$$

is the *standard composite average* of $(g_k)_{1\leqslant k\leqslant p}$ and $(L_k)_{1\leqslant k\leqslant p}$.

An alternative way to aggregate the functions $(g_k)_{1\leqslant k\leqslant p}$ and the linear operators $(L_k)_{1\leqslant k\leqslant p}$ is via the proximal comixture operation. This operation is derived from the proximal mixture operation recently introduced in [15] by duality, and it is further studied in [8, 18]. The main objective of the present paper is to propose the use of this aggregation process as an alternative to the standard composite average of Problem 1.2. This brings us to the following minimization model, which involves the conjugation operation of (2.1).

Problem 1.3. Let $\gamma \in \left]0,+\infty\right[$. Under Assumption 1.1, the task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + \big(\mathrm{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}\big)(x) + h(x), \tag{1.3}$$

where

$$\mathrm{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \left(\left(\sum_{k=1}^{p} \alpha_k \left(g_k^* + \gamma \mathscr{Q}_{\mathcal{G}_k}\right)^* \circ L_k\right)^* - \gamma \mathscr{Q}_{\mathcal{H}}\right)^* \tag{1.4}$$

is the *proximal comixture* of $(g_k)_{1\leqslant k\leqslant p}$ and $(L_k)_{1\leqslant k\leqslant p}$ with parameter γ.

Let us make some observations about these two formulations to motivate the latter.

- Both Problems 1.2 and 1.3 can be viewed as relaxations of the ideal feasibility problem

$$\text{find}\ \ x \in \mathcal{H} \quad \text{such that} \quad \begin{cases} x \in \operatorname{Argmin} f \\ (\forall k \in \{1, \dots, p\})\ L_k x \in \operatorname{Argmin} g_k \\ x \in \operatorname{Argmin} h, \end{cases} \tag{1.5}$$

which aims at imposing all the desired properties exactly and hence at minimizing all the loss functions simultaneously. However, the set

$$Z = \operatorname{Argmin} f \cap \left(\bigcap_{k=1}^{p} L_k^{-1}(\operatorname{Argmin} g_k) \right) \cap \operatorname{Argmin} h \tag{1.6}$$

of solutions to (1.5) is typically empty and, as will be seen in Remark 3.4(ii), both Problems 1.2 and 1.3 are relaxations of (1.5) in the sense that, if $Z \neq \varnothing$, then Z is the set of solutions of these two problems.

- The aggregation in Problem 1.2 may not be robust to perturbations. For instance, let us consider the special case when $f = 0$, $h = 0$, and, for every $k \in \{1, \dots, p\}$, $\mathcal{G}_k = \mathcal{H}$, $L_k = \mathrm{Id}_{\mathcal{H}}$, and g_k is the indicator function of a nonempty closed convex set $C_k \subset \mathcal{H}$. This reduces (1.1) to

$$\text{find}\ \ x \in \bigcap_{k=1}^{p} C_k, \tag{1.7}$$

which happens to coincide with (1.5) in this case. If the sets $(C_k)_{1\leqslant k\leqslant p}$ are not specified exactly, no solution may exist [10, 14]. However, it will follow from Remark 3.4(i) that the solutions to Problem 1.3 minimize the average squared distance to the sets. Such solutions are classical surrogates in inconsistent feasibility problems [14], which go back to Legendre's least-squares method for systems of linear equations [34]. They exist under mild conditions, such as the boundedness of one of the sets.

- Problem 1.2 involves a simple aggregation process by averaging. However, the nonsmooth function (1.2) has no explicit proximity operator, and solving (1.1) therefore calls for sophisticated proximal splitting methods to decompose the p functions $(g_k)_{1\leqslant k\leqslant p}$ and the p linear operators $(L_k)_{1\leqslant k\leqslant p}$ individually [16]. Overall, solving Problem 1.2 requires the splitting of $2p + 2$ terms, which can be expected to lead to algorithms that are slower and necessitate more memory storage than those that would split less terms. By contrast, the aggregated function (1.4) in Problem 1.3 is less intuitive but, as will be seen in Proposition 3.3(ii), its proximity operator can be computed explicitly in terms of those of the functions $(g_k)_{1\leqslant k\leqslant p}$ and of the linear operators $(L_k)_{1\leqslant k\leqslant p}$ as

$$\operatorname{prox}_{\gamma \operatorname{pcm}_\gamma (g_k, L_k)_{1\leqslant k\leqslant p}} = \mathrm{Id}_{\mathcal{H}} - \sum_{k=1}^{p} \alpha_k L_k^* \circ \left(\mathrm{Id}_{\mathcal{G}_k} - \operatorname{prox}_{\gamma g_k}\right) \circ L_k. \tag{1.8}$$

In turn, solving Problem 1.3 requires the splitting of only 3 terms.

- In the special case when $\sum_{k=1}^{p} \alpha_k = 1$ and, for every $k \in \{1, \dots, p\}$, $\mathcal{G}_k = \mathcal{H}$ and $L_k = \mathrm{Id}_{\mathcal{H}}$, the proximal comixture of (1.4) reduces to the proximal average

$$\operatorname{pav}_\gamma (g_k)_{1\leqslant k\leqslant p} = \left(\left(\sum_{k=1}^{p} \alpha_k \left(g_k^* + \gamma \mathcal{Q}_{\mathcal{H}} \right)^* \right)^* - \gamma \mathcal{Q}_{\mathcal{H}} \right)^*, \tag{1.9}$$

and (1.8) gives

$$\operatorname{prox}_{\gamma \operatorname{pav}_\gamma (g_k)_{1\leqslant k\leqslant p}} = \sum_{k=1}^{p} \alpha_k \operatorname{prox}_{\gamma g_k}. \tag{1.10}$$

This aggregation process has been implicitly introduced in [37], extensively studied in [6], and used in data analysis problems in [2, 13, 31, 36, 42, 48], where its benefits over the standard average

$$\operatorname{ave}(g_k)_{1\leqslant k\leqslant p} = \sum_{k=1}^{p} \alpha_k g_k, \tag{1.11}$$

are discussed.

- Formally, evaluating (1.4) at $\gamma = 0$ gives back (1.2). This connection will be made mathematically precise in Theorem 3.6, where we study the asymptotic behavior as $\gamma \downarrow 0$.

Notation and background are provided in Section 2. In Section 3, we study the mathematical properties of proximal comixtures and investigate connections between Problems 1.2 and 1.3. Section 4 is devoted to algorithms for solving Problems 1.2 and 1.3. Finally, Section 5 provides numerical illustrations of proximal comixture models in concrete signal restoration, image reconstruction, and linear regression problems.

§2. Notation and background

Throughout, $\mathcal{H}$ is a real Hilbert space with power set $2^{\mathcal{H}}$, identity operator $\mathrm{Id}_{\mathcal{H}}$, scalar product $\langle\cdot\mid\cdot\rangle_{\mathcal{H}}$, associated norm $\|\cdot\|_{\mathcal{H}}$, and quadratic kernel $\mathcal{Q}_{\mathcal{H}} = \|\cdot\|_{\mathcal{H}}^2/2$. For background on nonlinear analysis in Hilbert spaces, see [5].

Let $\varphi\colon \mathcal{H} \to [-\infty,+\infty]$. The conjugate of φ is

$$\varphi^*\colon \mathcal{H} \to [-\infty,+\infty]\colon x^* \mapsto \sup_{x\in\mathcal{H}} \big(\langle x \mid x^*\rangle_{\mathcal{H}} - \varphi(x)\big). \tag{2.1}$$

If $\inf_{y\in\mathcal{H}} \varphi(y) > -\infty$ and $\varepsilon \in [0,+\infty[$, the set of ε-minimizers of φ is

$$\varepsilon\text{–}\operatorname{Argmin}\varphi = \left\{x \in \mathcal{H} \mid \varphi(x) \leqslant \inf_{y\in\mathcal{H}} \varphi(y) + \varepsilon\right\}. \tag{2.2}$$

In particular, the set of minimizers of φ is $\operatorname{Argmin}\varphi = 0\text{–}\operatorname{Argmin}\varphi$. If $\operatorname{Argmin}\varphi$ is a singleton, its unique element is denoted by $\operatorname{argmin}_{x\in\mathcal{H}} \varphi(x)$. Moreover, φ is proper if $-\infty \notin \varphi(\mathcal{H})$ and $\operatorname{dom}\varphi = \{x \in \mathcal{H} \mid \varphi(x) < +\infty\} \neq \varnothing$. If φ is proper, its subdifferential is

$$\partial\varphi\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x \mid x^*\rangle_{\mathcal{H}} + \varphi(x) \leqslant \varphi(y)\big\}. \tag{2.3}$$

Let $\gamma \in \,]0,+\infty[$. Then the *(lower) Moreau envelope* of φ with parameter γ is

$$\operatorname{lenv}_\gamma \varphi\colon \mathcal{H} \to [-\infty,+\infty]\colon x \mapsto \inf_{y\in\mathcal{H}} \Big(\varphi(y) + \frac{1}{\gamma}\mathcal{Q}_{\mathcal{H}}(x-y)\Big), \tag{2.4}$$

and the *upper Moreau envelope* of φ with parameter γ is

$$\operatorname{uenv}_\gamma \varphi\colon \mathcal{H} \to [-\infty,+\infty]\colon x \mapsto \sup_{y\in\mathcal{H}} \Big(\varphi(y) - \frac{1}{\gamma}\mathcal{Q}_{\mathcal{H}}(x-y)\Big). \tag{2.5}$$

Given $\rho \in \mathbb{R}$, $\Gamma_\rho(\mathcal{H})$ denotes the class of proper lower semicontinuous functions $\varphi\colon \mathcal{H} \to]-\infty,+\infty]$ such that $\varphi + \rho\mathcal{Q}_\mathcal{H}$ is convex. The proximity operator of $\varphi \in \Gamma_0(\mathcal{H})$ is

$$\operatorname{prox}_\varphi\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y\in\mathcal{H}}{\operatorname{argmin}}\big(\varphi(y) + \mathcal{Q}_\mathcal{H}(x-y)\big) \tag{2.6}$$

and it is characterized by

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \operatorname{prox}_\varphi x \quad \Leftrightarrow \quad x - p \in \partial\varphi(p). \tag{2.7}$$

An operator $T\colon \mathcal{H} \to \mathcal{H}$ is γ-cocoercive if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle_\mathcal{H} \geqslant \gamma\|Tx - Ty\|_\mathcal{H}^2. \tag{2.8}$$

Let $C \subset \mathcal{H}$. Then the indicator function of C is denoted by ι_C and the distance function to C is denoted by d_C. If C is nonempty, closed, and convex, its projection operator is denoted by proj_C. The closed ball with center $x \in \mathcal{H}$ and radius $\rho \in]0,+\infty[$ is denoted by $B(x;\rho)$.

The following results will be useful.

Lemma 2.1. *Let $\varphi \in \Gamma_0(\mathcal{H})$ and let $\gamma \in]0,+\infty[$. Then the following hold:*

(i) $\operatorname{lenv}_\gamma \varphi = \big(\varphi^* + \gamma\mathcal{Q}_\mathcal{H}\big)^*$.

(ii) $\operatorname{uenv}_\gamma \varphi = \big(\varphi^* - \gamma\mathcal{Q}_\mathcal{H}\big)^*$.

Proof. Recall that, since $\varphi \in \Gamma_0(\mathcal{H})$, $\varphi^{**} = \varphi$ [5, Corollary 13.38].

(i): This follows from [5, Proposition 14.1].

(ii): Apply [29, Theorem 2.2] with $g = \varphi^*$ and $h = \gamma\mathcal{Q}_\mathcal{H}$. ☐

We recall the following result from [7, Lemma 3] (see also [40, Example 11.26(d)] for the finite-dimensional case).

Lemma 2.2. *Let $\varphi\colon \mathcal{H} \to]-\infty,+\infty]$ be proper and let $\gamma \in]0,+\infty[$. Then the following are equivalent:*

(i) $\operatorname{uenv}_\gamma (\operatorname{lenv}_\gamma \varphi) = \varphi$.

(ii) $\varphi \in \Gamma_{1/\gamma}(\mathcal{H})$.

Lemma 2.3. *Let $\varphi\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex, and let $\gamma \in]0,+\infty[$. Then the following are equivalent:*

(i) $\operatorname{lenv}_\gamma (\operatorname{uenv}_\gamma \varphi) = \varphi$.

(ii) $-\varphi \in \Gamma_{1/\gamma}(\mathcal{H})$.

(iii) $(\operatorname{uenv}_\gamma \varphi)^* = \varphi^* - \gamma\mathcal{Q}_\mathcal{H}$.

(iv) $\operatorname{uenv}_\gamma \varphi \in \Gamma_0(\mathcal{H})$, *$\varphi$ is Fréchet differentiable, and* $\operatorname{prox}_{\gamma(\operatorname{uenv}_\gamma \varphi)} = \mathrm{Id}_\mathcal{H} - \gamma\nabla\varphi$.

(v) *φ is Fréchet differentiable and $\nabla\varphi$ is γ-cocoercive.*

(vi) *φ is Fréchet differentiable and $\nabla\varphi$ is γ^{-1}-Lipschitzian.*

Proof. It follows from (2.4) and (2.5) that $\operatorname{lenv}_\gamma \varphi = -\operatorname{uenv}_\gamma(-\varphi)$. Therefore,

$$\operatorname{lenv}_\gamma\big(\operatorname{uenv}_\gamma \varphi\big) = -\operatorname{uenv}_\gamma\big(-\operatorname{uenv}_\gamma \varphi\big) = -\operatorname{uenv}_\gamma\big(\operatorname{lenv}_\gamma(-\varphi)\big). \tag{2.9}$$

(i)⇒(ii): By (2.9), $\operatorname{uenv}_\gamma(\operatorname{lenv}_\gamma(-\varphi)) = -\varphi$. Thus, Lemma 2.2 yields $-\varphi \in \Gamma_{1/\gamma}(\mathcal{H})$.

(ii)⇒(iii): By [5, Proposition 14.2], $\varphi^* \in \Gamma_{-\gamma}(\mathcal{H})$. Thus, the result follows from Lemma 2.1(ii) and [5, Corollary 13.38].

(iii)⇒(iv): Since $\varphi \in \Gamma_0(\mathcal{H})$, [5, Corollary 13.38] asserts that $\varphi^* \in \Gamma_0(\mathcal{H})$, which implies that $\operatorname{uenv}_\gamma \varphi$ is proper. In turn, it results from [5, Proposition 13.13] and Lemma 2.1(ii) that $\operatorname{uenv}_\gamma \varphi \in \Gamma_0(\mathcal{H})$. Thus, we derive from Lemma 2.1(i) and [5, Corollary 13.38] that

$$\operatorname{lenv}_\gamma(\operatorname{uenv}_\gamma \varphi) = \Big((\operatorname{uenv}_\gamma \varphi)^* + \gamma\mathcal{Q}_\mathcal{H}\Big)^* = \Big((\varphi^* - \gamma\mathcal{Q}_\mathcal{H}) + \gamma\mathcal{Q}_\mathcal{H}\Big)^* = \varphi^{**} = \varphi. \tag{2.10}$$

Hence, [5, Proposition 12.30] guarantees that $\varphi = \operatorname{lenv}_\gamma(\operatorname{uenv}_\gamma \varphi)$ is Fréchet differentiable and that $\nabla\varphi = \gamma^{-1}(\mathrm{Id}_\mathcal{H} - \operatorname{prox}_{\gamma(\operatorname{uenv}_\gamma \varphi)})$.

(iv)⇒(v): This follows from [5, Proposition 12.28].

(v)⇒(vi): This follows from (2.8) and the Cauchy–Schwarz inequality.

(vi)⇒(i): It follows from the equivalence (i)⇔(vi) in [5, Theorem 18.15] that $-\varphi \in \Gamma_{1/\gamma}(\mathcal{H})$. We therefore deduce from Lemma 2.2 that $\operatorname{uenv}_\gamma(\operatorname{lenv}_\gamma(-\varphi)) = -\varphi$. We conclude via (2.9). ☐

Remark 2.4. The functions $\operatorname{lenv}_\gamma \varphi$ and $\operatorname{uenv}_\gamma \varphi$ are, respectively, the infimal and the supremal convolutions of φ and $\mathcal{Q}_\mathcal{H}/\gamma$. In [30, Definition 2.4], $\operatorname{uenv}_\gamma \varphi$ is called the deconvolution of φ by $\mathcal{Q}_\mathcal{H}/\gamma$. Furthermore, when $\gamma < \rho$, the functions $\operatorname{lenv}_\gamma(\operatorname{uenv}_\rho \varphi)$ and $\operatorname{uenv}_\gamma(\operatorname{lenv}_\rho \varphi)$ are known as the *Lasry-Lions regularizations* of φ, which were introduced in [33] and further studied in [4, 7, 40, 43], whereas $\operatorname{lenv}_\gamma(\operatorname{uenv}_\gamma \varphi)$ is called the *proximal hull* of φ in [40, Example 1.44].

§3. Properties of proximal comixtures

Throughout this section, Assumption 1.1 is in force. Recall from (1.2) that the associated standard composite average is

$$\operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p} = \sum_{k=1}^{p} \alpha_k g_k \circ L_k, \tag{3.1}$$

and from (1.4) that the associated proximal comixture with parameter $\gamma \in \left]0,+\infty\right[$ is

$$\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \left(\left(\sum_{k=1}^{p} \alpha_k \Big(g_k^* + \gamma\mathcal{Q}_{\mathcal{G}_k}\Big)^* \circ L_k\right)^* - \gamma\mathcal{Q}_\mathcal{H}\right)^*. \tag{3.2}$$

We first provide reformulations of the proximal comixture of (3.2) in terms of the Moreau envelopes of (2.4) and (2.5) (see Fig. 1), and then in terms of the proximal average of (1.9) and of the following proximal cocomposition operation. These connections will not only shed new light on proximal comixtures but also play a role in forthcoming proofs.

Definition 3.1 ([15]). Let $\mathcal{G}$ be a real Hilbert space, let $g \in \Gamma_0(\mathcal{G})$, and let $L\colon \mathcal{H} \to \mathcal{G}$ be a bounded linear operator. The *proximal cocomposition* of g and L with parameter $\gamma \in \left]0,+\infty\right[$ is

$$L \overset{\gamma}{\blacklozenge} g = \left(\left(\big(g^* + \gamma\mathcal{Q}_\mathcal{G}\big)^* \circ L\right)^* - \gamma\mathcal{Q}_\mathcal{H}\right)^*. \tag{3.3}$$

Proposition 3.2. *Let $\gamma \in \left]0,+\infty\right[$. Then the following hold:*

(i) $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \operatorname{uenv}_\gamma \sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k$.

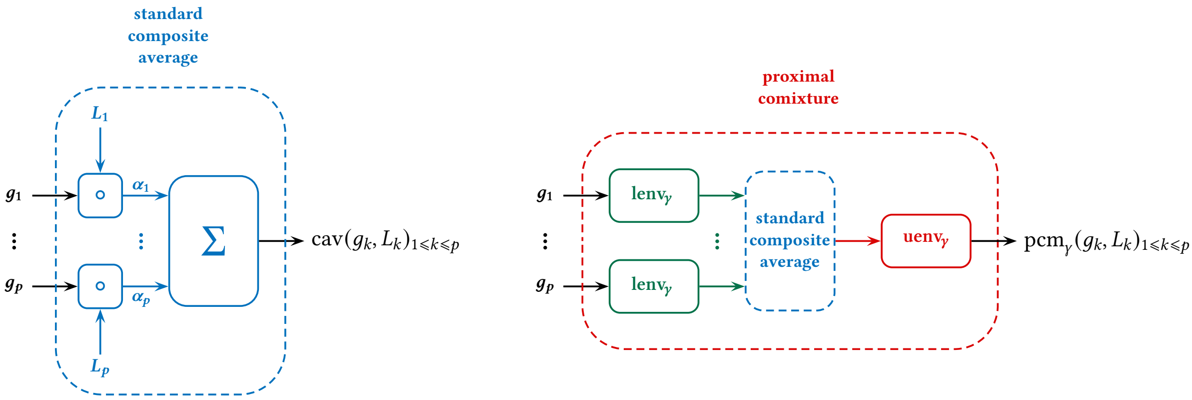


Figure 1: (left): Standard composite average (3.1). (right): Proximal comixture (3.2) in terms of the Moreau envelopes of (2.4) and (2.5) using Proposition 3.2(i).

(ii) *Suppose that* $\sum_{k=1}^p \alpha_k = 1$ *and* $\max_{1 \leqslant k \leqslant p} \|L_k\| \leqslant 1$. *Then* $\operatorname{pcm}_\gamma(g_k, L_k)_{1 \leqslant k \leqslant p} = \operatorname{pav}_\gamma(L_k \overset{\gamma}{\blacklozenge} g_k)_{1 \leqslant k \leqslant p}$.

(iii) *Let* $\mathcal{G}$ *be the standard product vector space* $\bigtimes_{k=1}^p \mathcal{G}_k$, *with generic element* $\boldsymbol{y} = (y_k)_{1 \leqslant k \leqslant p}$, *and equipped with the scalar product*

$$\langle \cdot \mid \cdot \rangle_{\mathcal{G}} \colon (\boldsymbol{y}, \boldsymbol{y}') \mapsto \sum_{k=1}^p \alpha_k \langle y_k \mid y'_k \rangle_{\mathcal{G}_k}, \tag{3.4}$$

and set

$$L \colon \mathcal{H} \to \mathcal{G} \colon x \mapsto (L_k x)_{1 \leqslant k \leqslant p} \quad \text{and} \quad g \colon \mathcal{G} \to \left]-\infty, +\infty\right] \colon \boldsymbol{y} \mapsto \sum_{k=1}^p \alpha_k g_k(y_k). \tag{3.5}$$

Then $\operatorname{pcm}_\gamma(g_k, L_k)_{1 \leqslant k \leqslant p} = L \overset{\gamma}{\blacklozenge} g$.

Proof. Set $\varphi = \sum_{k=1}^p \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k$. By virtue of [5, Proposition 12.30],

$$\nabla \varphi = \frac{1}{\gamma} \sum_{k=1}^p \alpha_k L_k^* \circ (\mathrm{Id}_{\mathcal{G}_k} - \operatorname{prox}_{\gamma g_k}) \circ L_k. \tag{3.6}$$

However, $\sum_{k=1}^p \alpha_k \|L_k\|^2 \leqslant 1$ by Assumption 1.1 and [5, Proposition 12.28] asserts that the operators $(\mathrm{Id}_{\mathcal{G}_k} - \operatorname{prox}_{\gamma g_k})_{1 \leqslant k \leqslant p}$ are nonexpansive. Therefore, (3.6) implies that $\nabla \varphi$ is γ^{-1}-Lipschitzian.

(i): We derive from (3.2) and Lemma 2.1(i)–(ii) that

$$\operatorname{pcm}_\gamma(g_k, L_k)_{1 \leqslant k \leqslant p} = \left(\varphi^* - \gamma \mathcal{Q}_{\mathcal{H}}\right)^* = \operatorname{uenv}_\gamma \varphi. \tag{3.7}$$

(ii): It follows from [18, Proposition 3.13(ii)] that $(\forall k \in \{1, \ldots, p\})$ $\operatorname{lenv}_\gamma (L_k \overset{\gamma}{\blacklozenge} g_k) = (\operatorname{lenv}_\gamma g_k) \circ L_k$.

Therefore, (i), Lemma 2.1(i)–(ii), and (1.9) yield

$$\begin{aligned}
\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} &= \operatorname{uenv}_\gamma \varphi \\
&= \operatorname{uenv}_\gamma \left(\sum_{k=1}^{p} \alpha_k \operatorname{lenv}_\gamma \left(L_k \overset{\gamma}{\blacklozenge} g_k\right)\right) \\
&= \left(\left(\sum_{k=1}^{p} \alpha_k \operatorname{lenv}_\gamma \left(L_k \overset{\gamma}{\blacklozenge} g_k\right)\right)^* - \gamma \mathscr{Q}_{\mathcal{H}}\right)^* \\
&= \operatorname{pav}_\gamma \left(L_k \overset{\gamma}{\blacklozenge} g_k\right)_{1\leqslant k\leqslant p}.
\end{aligned} \tag{3.8}$$

(iii): Since $\|\cdot\|_{\mathcal{G}}^2\colon \mathcal{G} \to \mathbb{R}\colon \boldsymbol{y} \mapsto \sum_{k=1}^{p} \alpha_k \|y_k\|_{\mathcal{G}_k}^2$, $\operatorname{lenv}_\gamma g\colon \mathcal{G} \to \mathbb{R}\colon \boldsymbol{y} \mapsto \sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k)(y_k)$. Therefore, $(\operatorname{lenv}_\gamma g) \circ L = \sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k$, and the result follows from (i), items (ii) and (i) in Lemma 2.1, and (3.3). ☐

Proposition 3.3. *Let $\gamma \in \left]0, +\infty\right[$. Then the following hold:*

(i) $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} \in \Gamma_0(\mathcal{H})$.

(ii) $\operatorname{prox}_{\gamma \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}} = \operatorname{Id}_{\mathcal{H}} - \sum_{k=1}^{p} \alpha_k L_k^* \circ (\operatorname{Id}_{\mathcal{G}_k} - \operatorname{prox}_{\gamma g_k}) \circ L_k$.

(iii) *Suppose that one of the following is satisfied:*

  (a) $\sum_{k=1}^{p} \alpha_k \|L_k\|^2 < 1$.

  (b) *For every $k \in \{1, \ldots, p\}$*, $\operatorname{dom} g_k = \mathcal{G}_k$.

  (c) $\sum_{k=1}^{p} \alpha_k = 1$, $\max_{1\leqslant k\leqslant p} \|L_k\| \leqslant 1$, *and there exists $\ell \in \{1, \ldots, p\}$ such that* $\operatorname{dom} g_\ell = \mathcal{G}_\ell$.

  Then $\operatorname{dom} \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \mathcal{H}$.

(iv) $\operatorname{lenv}_\gamma \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k$.

(v) $\operatorname{Argmin} \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \operatorname{Argmin} \sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k$.

(vi) *Suppose that $\bigcap_{k=1}^{p} L_k^{-1}(\operatorname{Argmin} g_k) \neq \varnothing$. Then*

$$\operatorname{Argmin} \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \operatorname{Argmin} \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p} = \bigcap_{k=1}^{p} L_k^{-1}(\operatorname{Argmin} g_k). \tag{3.9}$$

Proof. Define $\mathcal{G}$, L, and g as in (3.5), and note that

$$(\forall x \in \mathcal{H}) \quad \|Lx\|_{\mathcal{G}}^2 = \sum_{k=1}^{p} \alpha_k \|L_k x\|_{\mathcal{G}_k}^2 \leqslant \sum_{k=1}^{p} \alpha_k \|L_k\|^2 \|x\|_{\mathcal{H}}^2, \tag{3.10}$$

which yields $\|L\|^2 \leqslant \sum_{k=1}^{p} \alpha_k \|L_k\|^2 \leqslant 1$ by Assumption 1.1. Further, by Proposition 3.2(iii),

$$\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = L \overset{\gamma}{\blacklozenge} g. \tag{3.11}$$

(i): This follows from (3.11) and [18, Proposition 3.7(i)].

(ii): Note that $(\forall (y_k^*)_{1\leqslant k\leqslant p} \in \mathcal{G})\ L^*(y_k^*)_{1\leqslant k\leqslant p} = \sum_{k=1}^p \alpha_k L_k^* y_k^*$ and $\operatorname{prox}_{\gamma g}(y_k^*)_{1\leqslant k\leqslant p} = (\operatorname{prox}_{\gamma g_k} y_k^*)_{1\leqslant k\leqslant p}$. It therefore follows from (3.11) and [18, Proposition 3.10(ii)] that

$$\begin{aligned}
\operatorname{prox}_{\gamma \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}} &= \operatorname{prox}_{\gamma (L \overset{\gamma}{\bullet} g)} \\
&= \operatorname{Id}_{\mathcal{H}} - L^* \circ (\operatorname{Id}_{\mathcal{G}} - \operatorname{prox}_{\gamma g}) \circ L \\
&= \operatorname{Id}_{\mathcal{H}} - \sum_{k=1}^p \alpha_k L_k^* \circ (\operatorname{Id}_{\mathcal{G}_k} - \operatorname{prox}_{\gamma g_k}) \circ L_k. \qquad (3.12)
\end{aligned}$$

(iii)(a): In this case, $\|L\|^2 \leqslant \sum_{k=1}^p \alpha_k \|L_k\|^2 < 1$. Therefore, the assertion follows from (3.11) and [18, Proposition 3.2(iv)(a)].

(iii)(b): Since $\operatorname{dom} g = \bigtimes_{k=1}^p \operatorname{dom} g_k = \bigtimes_{k=1}^p \mathcal{G}_k = \mathcal{G}$, the assertion follows from (3.11) and [18, Proposition 3.2(iv)(b)].

(iii)(c): According to [18, Proposition 3.2(iv)(b)], $\operatorname{dom}(L_\ell \overset{\gamma}{\bullet} g_\ell) = \mathcal{H}$. Hence, we derive from Proposition 3.2(ii) and [15, Remark 5.11(v)] that

$$\operatorname{dom} \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \operatorname{dom} \operatorname{pav}_\gamma\big(L_k \overset{\gamma}{\bullet} g_k\big)_{1\leqslant k\leqslant p} = \sum_{k=1}^p \alpha_k \operatorname{dom}\big(L_k \overset{\gamma}{\bullet} g_k\big) = \mathcal{H}. \qquad (3.13)$$

(iv): Set $\varphi = \sum_{k=1}^p \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k$. As seen after (3.6), $\nabla\varphi$ is γ^{-1}-Lipschitzian. On the other hand, Proposition 3.2(i) asserts that $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \operatorname{uenv}_\gamma \varphi$. Altogether, appealing to the equivalence (i)⇔(vi) in Lemma 2.3, we conclude that $\operatorname{lenv}_\gamma \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \operatorname{lenv}_\gamma (\operatorname{uenv}_\gamma \varphi) = \varphi$.

(v): The result follows from (i), (iv), and the fact that the set of minimizers of a function in $\Gamma_0(\mathcal{H})$ coincides with that of its lower Moreau envelope [5, Proposition 17.5].

(vi): Since $\bigcap_{k=1}^p L_k^{-1}(\operatorname{Argmin} g_k) \neq \varnothing$, [5, Proposition 17.5] and (v) imply that

$$\begin{aligned}
\operatorname{Argmin} \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p} &= \bigcap_{k=1}^p L_k^{-1}(\operatorname{Argmin} g_k) \\
&= \bigcap_{k=1}^p L_k^{-1}(\operatorname{Argmin} \operatorname{lenv}_\gamma g_k) \\
&= \operatorname{Argmin} \sum_{k=1}^p \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k \\
&= \operatorname{Argmin} \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}, \qquad (3.14)
\end{aligned}$$

which completes the proof. ☐

Remark 3.4. Let us make a couple of observations about Proposition 3.3.

(i) Suppose that, in Assumption 1.1, $f = h = 0$ and, for every $k \in \{1, \ldots, p\}$, $g_k = \iota_{D_k}$ for some nonempty closed convex set $D_k \subset \mathcal{G}_k$. Then Problem 1.2 amounts to finding $x \in \mathcal{H}$ such that, for every $k \in \{1, \ldots, p\}$, $L_k x \in D_k$. On the other hand, it follows from Proposition 3.3(v) and [5, Example 12.21] that Problem 1.3 amounts to finding a minimizer of the least-squares function $x \mapsto \sum_{k=1}^p \alpha_k d_{D_k}^2(L_k x)$, which has been used in the relaxation of inconsistent feasibility problems [9, 14, 19].

(ii) In general, the solution sets of Problems 1.2 and 1.3 are distinct. However, it follows from Proposition 3.3(vi) that, if the set Z in (1.6) is nonempty, then it coincides with the solution set of both problems.

Next, we compare the standard composite average of $\operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$ of (3.1) with the proximal mixture $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}$ of (3.2). Let us start with some basic inequalities.

Proposition 3.5. *Let $\gamma \in \left]0, +\infty\right[$ and*

$$\varphi\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \inf_{\substack{1\leqslant k\leqslant p\\ y_k^*\in\partial g_k(L_k x)}} \left(\sum_{k=1}^{p} \alpha_k \mathcal{Q}_{\mathcal{G}_k}(y_k^*) - \mathcal{Q}_{\mathcal{H}}\left(\sum_{k=1}^{p} \alpha_k L_k^* y_k^*\right)\right). \tag{3.15}$$

Then $\sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k \leqslant \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} \leqslant \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p} \leqslant \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} + \gamma\varphi$.

Proof. Define $\mathcal{G}$, L, and g as in (3.5), and recall from Proposition 3.2(iii) that $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = L \overset{\gamma}{\blacklozenge} g$. Further, $g \circ L = \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$, and as seen after (3.10), $\|L\| \leqslant 1$. It follows from [18, Proposition 3.20(ii)] that

$$\sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_\gamma g_k) \circ L_k = (\operatorname{lenv}_\gamma g) \circ L \leqslant L \overset{\gamma}{\blacklozenge} g \leqslant g \circ L, \tag{3.16}$$

where $L \overset{\gamma}{\blacklozenge} g = \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}$ and $g\circ L = \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$. It remains to show the rightmost inequality. Let $x \in \mathcal{H}$. The result is clear if $x \notin \operatorname{dom}\varphi$. Now, suppose that $x \in \operatorname{dom}\varphi$ and let $\boldsymbol{y}^* = (y_k^*)_{1\leqslant k\leqslant p} \in \bigtimes_{k=1}^{p} \partial g_k(L_k x)$. Then (3.4)–(3.5) yield $L^*\boldsymbol{y}^* = \sum_{k=1}^{p} \alpha_k L_k^* y_k^*$ and $\mathcal{Q}_{\mathcal{G}}(\boldsymbol{y}^*) = \sum_{k=1}^{p} \alpha_k \mathcal{Q}_{\mathcal{G}_k}(y_k^*)$. On the other hand, by [5, Proposition 16.9], $\boldsymbol{y}^* \in \bigtimes_{k=1}^{p} \partial g_k(L_k x) = \partial g(Lx)$ and we therefore deduce from [18, Proposition 3.23(i)] that

$$0 \leqslant \big(\operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}\big)(x) - \big(\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}\big)(x) \leqslant \gamma\big(\mathcal{Q}_{\mathcal{G}}(\boldsymbol{y}^*) - \mathcal{Q}_{\mathcal{H}}(L^*\boldsymbol{y}^*)\big). \tag{3.17}$$

Hence, taking the infimum over $\boldsymbol{y}^* \in \bigtimes_{k=1}^{p} \partial g_k(L_k x)$ yields the claim. ☐

The following result establishes asymptotic relations between standard composite averages and proximal comixtures.

Theorem 3.6. *Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, +\infty\right[$ such that $\gamma_n \downarrow 0$ and let $x \in \mathcal{H}$. For brevity, write $\mathfrak{c} = \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$ and $(\forall n \in \mathbb{N})$ $\mathfrak{m}_{\gamma_n} = \operatorname{pcm}_{\gamma_n}(g_k, L_k)_{1\leqslant k\leqslant p}$. Then the following hold:*

(i) $\mathfrak{m}_{\gamma_n}(x) \uparrow \mathfrak{c}(x)$.

(ii) *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $x_n \rightharpoonup x$. Then $\mathfrak{c}(x) \leqslant \underline{\lim}\, \mathfrak{m}_{\gamma_n}(x_n)$.*

(iii) *Suppose that $\mathfrak{c}$ is proper and let $\gamma \in \left]0, +\infty\right[$. Then $\operatorname{prox}_{\gamma\mathfrak{m}_{\gamma_n}} x \to \operatorname{prox}_{\gamma\mathfrak{c}} x$.*

(iv) *Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, +\infty\right[$ such that $\varepsilon_n \downarrow 0$, and suppose that there exists a bounded sequence $(z_n)_{n\in\mathbb{N}}$ such that $(\forall n \in \mathbb{N})$ $z_n \in \varepsilon_n$–$\operatorname{Argmin}(f + \mathfrak{m}_{\gamma_n} + h)$. Then the following hold:*

   (a) $\inf_{x\in\mathcal{H}}(f(x) + \mathfrak{m}_{\gamma_n}(x) + h(x)) \uparrow \min_{x\in\mathcal{H}}(f(x) + \mathfrak{c}(x) + h(x))$.

   (b) *Every weak sequential cluster point of $(z_n)_{n\in\mathbb{N}}$ is in $\operatorname{Argmin}(f + \mathfrak{c} + h)$.*

Proof. (i): Define L and g as in (3.5), and recall from Proposition 3.2(iii) that $(\forall n \in \mathbb{N})\ \mathfrak{m}_{\gamma_n} = L \overset{\gamma_n}{\blacklozenge} g$. By items (ii) and (iv) in [18, Theorem 3.29], the function $\mathbb{N} \to \left]-\infty,+\infty\right] \colon n \mapsto \mathfrak{m}_{\gamma_n}(x)$ is increasing and $\lim_{n\to+\infty} \mathfrak{m}_{\gamma_n}(x) = g(Lx) = \mathfrak{c}(x)$.

(ii): The weak continuity of bounded linear operators [5, Lemma 2.41] yields $(\forall k \in \{1,\ldots,p\})$ $L_k x_n \rightharpoonup L_k x$. Thus, [38, Proposition 2.2(d)] implies that

$$(\forall k \in \{1,\ldots,p\}) \quad g_k(L_k x) \leqslant \underline{\lim}(\operatorname{lenv}_{\gamma_n} g_k)(L_k x_n). \tag{3.18}$$

On the other hand, recall that $\mathfrak{c} = \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$. Therefore, it follows from (3.18), [5, Lemma 1.16], and Proposition 3.5 that

$$\mathfrak{c}(x) \leqslant \sum_{k=1}^{p} \alpha_k \underline{\lim}(\operatorname{lenv}_{\gamma_n} g_k)(L_k x_n) \leqslant \underline{\lim} \sum_{k=1}^{p} \alpha_k (\operatorname{lenv}_{\gamma_n} g_k)(L_k x_n) \leqslant \underline{\lim}\, \mathfrak{m}_{\gamma_n}(x_n). \tag{3.19}$$

(iii): Recall from Proposition 3.3(i) that $(\forall n \in \mathbb{N})\ \mathfrak{m}_{\gamma_n} \in \Gamma_0(\mathcal{H})$. Therefore, the result follows from (i), (ii), the equivalence (i)⇔(iii) in [3, Proposition 3.19], and the equivalence (i)⇔(ii) in [3, Theorem 3.26].

(iv): By Proposition 3.3(i) and (i), $(\mathfrak{m}_{\gamma_n})_{n\in\mathbb{N}}$ is an increasing sequence of functions in $\Gamma_0(\mathcal{H})$ with $\sup_{n\in\mathbb{N}} \mathfrak{m}_{\gamma_n} = \mathfrak{c}$. Therefore, the equivalence (iii)⇔(iv) in [5, Theorem 9.1] implies that $(f + \mathfrak{m}_{\gamma_n} + h)_{n\in\mathbb{N}}$ is an increasing sequence of weakly lower semicontinuous convex functions with supremum $f + \mathfrak{c} + h$. Since $(z_n)_{n\in\mathbb{N}}$ is bounded, [5, Lemma 2.45] asserts that $(z_n)_{n\in\mathbb{N}}$ admits a weakly convergent subsequence. Hence, the results follow from [3, Proposition 2.42] applied to the weak topology. ☐

The following result focuses on the case in which the functions $(g_k)_{1\leqslant k\leqslant p}$ are Lipschitzian.

Proposition 3.7. *Assume that, for every $k \in \{1,\ldots,p\}$, g_k is real-valued and μ_k-Lipschitzian for some $\mu_k \in \left]0,+\infty\right[$. Set $\mu = \sum_{k=1}^{p} \alpha_k \mu_k \|L_k\|$, set $\vartheta = (1/2)\sum_{k=1}^{p} \alpha_k \mu_k^2$, let $x \in \mathcal{H}$, and let $\gamma \in \left]0,+\infty\right[$. Then the following hold:*

(i) $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}$ *is μ-Lipschitzian.*

(ii) $0 \leqslant \operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p} - \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} \leqslant \gamma\vartheta$.

(iii) *There exists $r \in B(0; 2\gamma\mu)$ such that* $\operatorname{prox}_{\gamma \operatorname{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}} x = \operatorname{prox}_{\gamma \operatorname{cav}(g_k,L_k)_{1\leqslant k\leqslant p}}(x + r)$.

(iv) $\|\operatorname{prox}_{\gamma \operatorname{cav}(g_k,L_k)_{1\leqslant k\leqslant p}} x - \operatorname{prox}_{\gamma \operatorname{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}} x\| \leqslant \gamma \min\{2\mu, \sqrt{2\vartheta}\}$.

(v) *Let $\sigma \in \left]0,+\infty\right[$ and assume that $h \in \Gamma_{-\sigma}(\mathcal{H})$. Then*

$$\Big\|\operatorname{argmin}\big(\operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p} + h\big) - \operatorname{argmin}\big(\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} + h\big)\Big\|_{\mathcal{H}} \leqslant \frac{2\mu}{\sigma}. \tag{3.20}$$

Proof. According to [5, Corollary 17.19], a lower semicontinuous convex function $\varphi \colon \mathcal{H} \to \mathbb{R}$ is μ-Lipschitzian if and only if $\operatorname{range} \partial\varphi \subset B(0;\mu)$. Thus, the Lipschitz continuity of the functions $(g_k)_{1\leqslant k\leqslant p}$ implies that

$$(\forall k \in \{1,\ldots,p\}) \quad \operatorname{range} \partial g_k \subset B(0;\mu_k). \tag{3.21}$$

Further, by [5, Proposition 16.27], $(\forall k \in \{1,\ldots,p\})\ \operatorname{dom} \partial g_k = \mathcal{G}_k$.

(i): By virtue of Proposition 3.3(iii)(b), $\operatorname{dom} \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \mathcal{H}$ and therefore [5, Proposition 16.27] yields $\operatorname{dom} \partial \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \mathcal{H}$. Now, let $u \in \operatorname{range} \partial \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}$. Then there exists $x \in \mathcal{H}$ such that $u \in (\partial \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p})(x)$ which, by (2.7), is equivalent to

$\operatorname{prox}_{\gamma\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}}(x+\gamma u)=x$. Now, set $(\forall k\in\{1,\ldots,p\})$ $y_k^*=L_k(x+\gamma u)-\operatorname{prox}_{\gamma g_k}(L_k(x+\gamma u))$. By Proposition 3.3(ii),

$$x+\gamma u-\sum_{k=1}^p\alpha_k L_k^*y_k^*=x. \tag{3.22}$$

Further, it follows from (2.7) and (3.21) that

$$(\forall k\in\{1,\ldots,p\})\quad \frac{1}{\gamma}y_k^*\in\partial g_k\big(L_k(x+\gamma u)-y_k^*\big)\subset\operatorname{range}\partial g_k\subset B(0;\mu_k). \tag{3.23}$$

Therefore, it follows from (3.22) and (3.23) that

$$\|u\|_{\mathcal{H}}=\left\|\frac{1}{\gamma}\sum_{k=1}^p\alpha_kL_k^*y_k^*\right\|_{\mathcal{H}}\leqslant\sum_{k=1}^p\alpha_k\|L_k\|\left\|\frac{1}{\gamma}y_k^*\right\|_{\mathcal{G}_k}\leqslant\sum_{k=1}^p\alpha_k\|L_k\|\mu_k=\mu. \tag{3.24}$$

Altogether, $\operatorname{range}\partial\,\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}\subset B(0;\mu)$, which yields the conclusion via [5, Corollary 17.19].

(ii): Recall that $(\forall k\in\{1,\ldots,p\})$ $\operatorname{dom}\partial g_k=\mathcal{G}_k$. For every $k\in\{1,\ldots,p\}$, let $y_k^*\in\partial g_k(L_kx)$. Thus, it follows from Proposition 3.5 and (3.21) that

$$0\leqslant\big(\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p}\big)(x)-\big(\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}\big)(x)\leqslant\gamma\sum_{k=1}^p\alpha_k\mathcal{Q}_{\mathcal{G}_k}(y_k^*)\leqslant\frac{\gamma}{2}\sum_{k=1}^p\alpha_k\mu_k^2=\gamma\vartheta. \tag{3.25}$$

(iii): Set $u_\gamma=\operatorname{prox}_{\gamma\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}}x$. Then, by (2.7), $x-u_\gamma\in\gamma(\partial\,\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p})(u_\gamma)$. On the other hand, since $\operatorname{dom}\partial\,\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p}=\mathcal{H}$, there exists $r\in\mathcal{H}$ such that $x+r-u_\gamma\in\gamma(\partial\,\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p})(u_\gamma)$. Thus, by (2.7), $\operatorname{prox}_{\gamma\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p}}(x+r)=u_\gamma$. Further, by (i) and [5, Corollary 17.19],

$$\operatorname{range}\partial\,\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p}\subset B(0;\mu)\quad\text{and}\quad\operatorname{range}\partial\,\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}\subset B(0;\mu). \tag{3.26}$$

Therefore, $r=(x+r-u_\gamma)-(x-u_\gamma)\in\gamma B(0;\mu)+\gamma B(0;\mu)=B(0;2\gamma\mu)$.

(iv): Set $u=\operatorname{prox}_{\gamma\,\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p}}x$, $u_\gamma=\operatorname{prox}_{\gamma\mathrm{pcm}_\gamma(g_k,L_k)_{1\leqslant k\leqslant p}}x$, and

$$(\forall k\in\{1,\ldots,p\})\quad y_k^*=L_kx-\operatorname{prox}_{\gamma g_k}(L_kx). \tag{3.27}$$

Further, set $z=\sum_{k=1}^p\alpha_kL_k^*y_k^*$. Then, Proposition 3.3(ii) yields $x-z=u_\gamma$, whereas (2.7), (3.27), and (3.21) imply that

$$(\forall k\in\{1,\ldots,p\})\quad y_k^*\in\gamma\partial g_k(L_kx-y_k^*)\subset B(0;\gamma\mu_k). \tag{3.28}$$

By [5, Corollaries 16.48(iii) and 16.53(i)], $x-u\in\gamma(\partial\,\mathrm{cav}(g_k,L_k)_{1\leqslant k\leqslant p})(u)=\gamma\sum_{k=1}^p\alpha_kL_k^*(\partial g_k(L_ku))$. Thus, for every $k\in\{1,\ldots,p\}$, there exists $w_k^*\in\gamma\partial g_k(L_ku)$ such that $x-u=\sum_{k=1}^p\alpha_kL_k^*w_k^*$. It follows from (3.28) and the monotonicity of the subdifferential operators $(\gamma\partial g_k)_{1\leqslant k\leqslant p}$ [5, Theorem 20.25] that

$$(\forall k\in\{1,\ldots,p\})\quad\langle L_kx-y_k^*-L_ku\mid y_k^*-w_k^*\rangle_{\mathcal{G}_k}\geqslant 0. \tag{3.29}$$

Since $\sum_{k=1}^p\alpha_k\|L_k\|^2\leqslant 1$ by Assumption 1.1, (3.28) implies that

$$\sum_{k=1}^p\alpha_k\|y_k^*-L_kz\|_{\mathcal{G}_k}^2=\sum_{k=1}^p\alpha_k\|y_k^*\|_{\mathcal{G}_k}^2-2\|z\|_{\mathcal{H}}^2+\sum_{k=1}^p\alpha_k\|L_kz\|_{\mathcal{G}_k}^2\leqslant\sum_{k=1}^p\alpha_k\|y_k^*\|_{\mathcal{G}_k}^2\leqslant 2\gamma^2\vartheta, \tag{3.30}$$

whereas the Cauchy–Schwarz inequality yields

$$\|z\|_{\mathcal{H}}^2 \leqslant \left(\sum_{k=1}^{p} \alpha_k \|L_k\| \|y_k^*\|_{\mathcal{G}_k}\right)^2 \leqslant \left(\sum_{k=1}^{p} \alpha_k \|L_k\|^2\right)\left(\sum_{k=1}^{p} \alpha_k \|y_k^*\|_{\mathcal{G}_k}^2\right) \leqslant \sum_{k=1}^{p} \alpha_k \|y_k^*\|_{\mathcal{G}_k}^2. \tag{3.31}$$

Altogether, we deduce from $x = z + u_\gamma$, $\sum_{k=1}^{p} \alpha_k L_k^* y_k^* = z$, $\sum_{k=1}^{p} \alpha_k L_k^* w_k^* = x - u$, (3.29), the Cauchy–Schwarz inequality, (3.30), and (3.31) that

$$\begin{aligned}
0 &\leqslant \sum_{k=1}^{p} \alpha_k \langle L_k u_\gamma - L_k u + L_k z - y_k^* \mid y_k^* - w_k^* \rangle_{\mathcal{G}_k} \\
&= \langle u_\gamma - u \mid (x - u_\gamma) - (x - u) \rangle_{\mathcal{H}} + \sum_{k=1}^{p} \alpha_k \langle L_k z - y_k^* \mid y_k^* - w_k^* \rangle_{\mathcal{G}_k} \\
&= -\|u - u_\gamma\|_{\mathcal{H}}^2 + \sum_{k=1}^{p} \alpha_k \langle y_k^* - L_k z \mid w_k^* \rangle_{\mathcal{G}_k} - \sum_{k=1}^{p} \alpha_k \langle y_k^* - L_k z \mid y_k^* \rangle_{\mathcal{G}_k} \\
&\leqslant -\|u - u_\gamma\|_{\mathcal{H}}^2 + \sqrt{\sum_{k=1}^{p} \alpha_k \|y_k^* - L_k z\|_{\mathcal{G}_k}^2}\sqrt{\sum_{k=1}^{p} \alpha_k \|w_k^*\|_{\mathcal{G}_k}^2} + \left(\|z\|_{\mathcal{H}}^2 - \sum_{k=1}^{p} \alpha_k \|y_k^*\|_{\mathcal{G}_k}^2\right) \\
&\leqslant -\|u - u_\gamma\|_{\mathcal{H}}^2 + \sqrt{2\gamma^2 \vartheta}\sqrt{\gamma^2 \sum_{k=1}^{p} \alpha_k \mu_k^2} \\
&= -\|u - u_\gamma\|_{\mathcal{H}}^2 + 2\gamma^2 \vartheta.
\end{aligned} \tag{3.32}$$

However, by (iii), there exists $r \in B(0; 2\gamma\mu)$ such that $\operatorname{prox}_{\gamma \operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p}} x = \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} (x+r)$. Hence, the nonexpansiveness of the proximity operator [5, Proposition 12.28] implies that

$$\|u - u_\gamma\|_{\mathcal{H}} = \left\| \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} x - \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} (x + r) \right\|_{\mathcal{H}} \leqslant \|r\|_{\mathcal{H}} \leqslant 2\gamma\mu. \tag{3.33}$$

Finally, the result follows from (3.32) and (3.33).

(v): Note that $\operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p} + h \in \Gamma_{-\sigma}(\mathcal{H})$ and, by Proposition 3.3(i), also $\operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p} + h \in \Gamma_{-\sigma}(\mathcal{H})$. Thus, [5, Corollary 11.17] implies that $\operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p} + h$ and $\operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p} + h$ admit unique minimizers. Set $x = \operatorname{argmin}(\operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p} + h)$ and $x_\gamma = \operatorname{argmin}(\operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p} + h)$. Then the equivalence (i)⇔(viii) in [5, Corollary 27.3] yields

$$x = \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x - \gamma \nabla h(x)\big) \quad \text{and} \quad x_\gamma = \operatorname{prox}_{\gamma \operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x_\gamma - \gamma \nabla h(x_\gamma)\big). \tag{3.34}$$

On the other hand, (iii) asserts that there exists $r \in B(0; 2\gamma\mu)$ such that

$$\operatorname{prox}_{\gamma \operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x - \gamma \nabla h(x)\big) = \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x - \gamma \nabla h(x) + r\big). \tag{3.35}$$

Altogether, (3.34), (3.35), and the firm nonexpansiveness of $\operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}}$ [5, Proposition 12.28] yield

$$\begin{aligned}
\|x - x_\gamma\|_{\mathcal{H}}^2 &= \left\| \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x - \gamma \nabla h(x)\big) - \operatorname{prox}_{\gamma \operatorname{pcm}_\gamma (g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x_\gamma - \gamma \nabla h(x_\gamma)\big) \right\|_{\mathcal{H}}^2 \\
&= \left\| \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x - \gamma \nabla h(x)\big) - \operatorname{prox}_{\gamma \operatorname{cav}(g_k, L_k)_{1 \leqslant k \leqslant p}} \big(x_\gamma - \gamma \nabla h(x_\gamma) + r\big) \right\|_{\mathcal{H}}^2 \\
&\leqslant \langle x - \gamma \nabla h(x) - x_\gamma + \gamma \nabla h(x_\gamma) - r \mid x - x_\gamma \rangle_{\mathcal{H}} \\
&= \|x - x_\gamma\|_{\mathcal{H}}^2 - \gamma \langle \nabla h(x) - \nabla h(x_\gamma) \mid x - x_\gamma \rangle_{\mathcal{H}} - \langle r \mid x - x_\gamma \rangle_{\mathcal{H}}.
\end{aligned} \tag{3.36}$$

Therefore, it follows from [5, Example 22.4(iv)], the Cauchy–Schwarz inequality, and the inequality $\|r\|_{\mathcal{H}} \leqslant 2\gamma\mu$ that

$$\gamma\sigma\|x - x_\gamma\|_{\mathcal{H}}^2 \leqslant \gamma\langle \nabla h(x) - \nabla h(x_\gamma) \mid x - x_\gamma\rangle_{\mathcal{H}} \leqslant 2\gamma\mu\|x - x_\gamma\|_{\mathcal{H}}, \tag{3.37}$$

which implies (3.20). ☐

As a special case of the above results, we recover some properties of the proximal average [6, 44, 48].

Remark 3.8 (proximal average). Suppose that $\sum_{k=1}^p \alpha_k = 1$ and that, for every $k \in \{1, \ldots, p\}$, $\mathcal{G}_k = \mathcal{H}$ and $L_k = \mathrm{Id}_{\mathcal{H}}$. Then $\mathrm{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} = \mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p}$ is the proximal average of (1.9), while $\mathrm{cav}(g_k, L_k)_{1\leqslant k\leqslant p} = \mathrm{ave}(g_k)_{1\leqslant k\leqslant p}$ is the standard average of (1.11). In this context, we recover the following results:

(i) Proposition 3.3(i) yields $\mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p} \in \Gamma_0(\mathcal{H})$ (see [6, Corollary 5.2]).

(ii) Proposition 3.3(ii) yields $\mathrm{prox}_{\gamma \mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p}} = \sum_{k=1}^p \alpha_k \,\mathrm{prox}_{\gamma g_k}$ (see [6, Theorem 6.7]).

(iii) Proposition 3.3(iv) yields $\mathrm{lenv}_\gamma \mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p} = \sum_{k=1}^p \alpha_k \mathrm{lenv}_\gamma\, g_k$ (see [6, Theorem 6.2(i)]).

(iv) Proposition 3.5 yields $\mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p} \leqslant \mathrm{ave}$ (see [6, Theorem 5.4]).

(v) Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\gamma_n \downarrow 0$.

  (a) Theorem 3.6(i) yields $\mathrm{pav}_{\gamma_n}(g_k)_{1\leqslant k\leqslant p} \uparrow \mathrm{ave}(g_k)_{1\leqslant k\leqslant p}$ (see [6, Theorem 8.5]).

  (b) Theorem 3.6(i)–(ii) imply that $(\mathrm{pav}_{\gamma_n}(g_k)_{1\leqslant k\leqslant p})_{n\in\mathbb{N}}$ epi-converges to $\mathrm{ave}(g_k)_{1\leqslant k\leqslant p}$ (see [6, Corollary 9.6]).

(vi) Suppose that, for every $k \in \{1, \ldots, p\}$, g_k is real-valued and μ_k-Lipschitzian for some $\mu_k \in]0, +\infty[$. Set $\mu = \sum_{k=1}^p \alpha_k\mu_k$ and $\vartheta = (1/2)\sum_{k=1}^p \alpha_k\mu_k^2$.

  (a) Proposition 3.7(ii) yields $0 \leqslant \mathrm{ave}(g_k)_{1\leqslant k\leqslant p} - \mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p} \leqslant \gamma\vartheta$ (see [48, Proposition 4]).

  (b) Proposition 3.7(iv) yields $\|\,\mathrm{prox}_{\gamma\mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p}} - \mathrm{prox}_{\gamma\mathrm{ave}(g_k)_{1\leqslant k\leqslant p}}\|_{\mathcal{H}} \leqslant 2\gamma\mu$ (see [44, p. 851]).

In addition, we obtain the following new properties of the proximal average.

(vii) Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\gamma_n \downarrow 0$. It follows from Theorem 3.6(iii) that $\mathrm{prox}_{\gamma\mathrm{pav}_{\gamma_n}(g_k)_{1\leqslant k\leqslant p}} x \to \mathrm{prox}_{\gamma\mathrm{ave}(g_k)_{1\leqslant k\leqslant p}} x$.

(viii) Suppose that, for every $k \in \{1, \ldots, p\}$, g_k is real-valued and μ_k-Lipschitzian for some $\mu_k \in]0, +\infty[$. Set $\mu = \sum_{k=1}^p \alpha_k\mu_k$. Then, by Proposition 3.7(iii), there exists $r \in B(0; 2\gamma\mu)$ such that $\mathrm{prox}_{\gamma\mathrm{pav}_\gamma(g_k)_{1\leqslant k\leqslant p}} x = \mathrm{prox}_{\gamma\mathrm{ave}(g_k)_{1\leqslant k\leqslant p}}(x + r)$.

Generalization of Problems 1.2 and 1.3 to arbitrary families of functions and linear operators can be formulated as follows.

Remark 3.9 (integral proximal comixture). Let $(\Omega, \mathcal{F}, \mu)$ be a complete σ-finite measure space and let H be a separable real Hilbert space. For every $\omega \in \Omega$, let G_ω be a real Hilbert space, let $\mathsf{g}_\omega \in \Gamma_0(\mathsf{G}_\omega)$, and let $\mathsf{L}_\omega \colon \mathsf{H} \to \mathsf{G}_\omega$ be a bounded linear operator. Assume that $0 < \int_\Omega \|\mathsf{L}_\omega\|^2\mu(d\omega) \leqslant 1$. Under mild assumptions, the *integral proximal comixture* of $(\mathsf{g}_\omega)_{\omega\in\Omega}$ and $(\mathsf{L}_\omega)_{\omega\in\Omega}$ with parameter $\gamma \in]0, +\infty[$ is [8, Definition 4.2]

$$\overset{\bullet}{\mathsf{M}}_\gamma(\mathsf{g}_\omega, \mathsf{L}_\omega)_{\omega\in\Omega} = \left(\varphi^* - \frac{\gamma}{2}\|\cdot\|_{\mathsf{H}}^2\right)^*, \quad \text{where} \quad (\forall \mathsf{x} \in \mathsf{H}) \quad \varphi(\mathsf{x}) = \int_\Omega (\mathrm{lenv}_\gamma\, \mathsf{g}_\omega)(\mathsf{L}_\omega \mathsf{x})\mu(d\omega). \tag{3.38}$$

Now let $\mathsf{f} \in \Gamma_0(\mathsf{H})$ and let $\mathsf{h}\colon \mathsf{H} \to \mathbb{R}$ be a convex and differentiable function with Lipschitzian gradient. Then a generalization of Problem 1.2 is

$$\underset{\mathsf{x}\in\mathsf{H}}{\text{minimize}}\;\; \mathsf{f}(\mathsf{x}) + \int_{\Omega} \mathsf{g}_\omega(\mathsf{L}_\omega \mathsf{x})\,\mu(d\omega) + \mathsf{h}(\mathsf{x}) \tag{3.39}$$

and a generalization of Problem 1.3 is

$$\underset{\mathsf{x}\in\mathsf{H}}{\text{minimize}}\;\; \mathsf{f}(\mathsf{x}) + \Big(\overset{\bullet}{\mathsf{M}}_\gamma(\mathsf{g}_\omega, \mathsf{L}_\omega)_{\omega\in\Omega}\Big)(\mathsf{x}) + \mathsf{h}(\mathsf{x}). \tag{3.40}$$

Indeed, we recover (1.1) from (3.39) and (1.3) from (3.40) by taking $\Omega = \{1,\ldots,p\}$ and $\mathcal{F} = 2^{\Omega}$, and setting $(\forall k \in \{1,\ldots,p\})$ $\mu(\{k\}) = \alpha_k$. Most of the results of this section extend to this abstract framework using the tools of [8, 18]. If μ is a probability measure, we can regard (1.1) and (1.3) as empirical versions of (3.39) and (3.40), respectively.

§4. Algorithms

Several splitting methods are available to solve Problems 1.2 and 1.3 [16]. For solving Problem 1.2, methods that require the proximity operator of $\operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$ are ruled out as this operator is not tractable. We have therefore considered those of [16, 21, 24, 46, 47] and found that, overall, the following method [47] (see also [26]), which is based on [27], performed the best in our experiments.

Proposition 4.1. *Consider the setting of Problem 1.2 and assume that*

$$0 \in \operatorname{range}\Big(\partial f + \sum_{k=1}^{p} \alpha_k L_k^* \circ (\partial g_k) \circ L_k + \nabla h\Big). \tag{4.1}$$

Let $\eta \in \left]0, 2\beta\right[$ and $\sigma \in \left]0, 1/(\eta \sum_{k=1}^{p} \alpha_k \|L_k\|^2)\right[$. Let $q_0 \in \mathcal{H}$ and, for every $k \in \{1,\ldots,p\}$, let $u_{k,0}^* \in \mathcal{G}_k$. Iterate

$$\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_n = \operatorname{prox}_{\eta f}\Big(\eta\big(q_n - \sum_{k=1}^{p} \alpha_k L_k^* u_{k,n}^*\big)\Big)\\
q_{n+1} = \dfrac{1}{\eta} x_n - \nabla h(x_n)\\
\text{for } k = 1,\ldots,p\\
\left\lfloor
\begin{array}{l}
r_{k,n} = u_{k,n}^* + \sigma L_k\big(x_n + \eta(q_{n+1} - q_n)\big)\\
u_{k,n+1}^* = r_{k,n} - \sigma \operatorname{prox}_{g_k/\sigma}(r_{k,n}/\sigma).
\end{array}
\right.
\end{array}
\right.
\end{array} \tag{4.2}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution to Problem 1.2.

Proof. Apply [26, Theorem 1] in the Hilbert space $\mathcal{G}$ of (3.4), with g and L defined as in (3.5). ☐

We now turn to Problem 1.3, for which the method proposed in [27, Section 3.1] can be implemented directly. Given $y_0 \in \mathcal{H}$, this algorithm iterates

$$\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_n = \operatorname{prox}_{\gamma \operatorname{pcm}_\gamma (g_k, L_k)_{1\leqslant k\leqslant p}} y_n\\
z_n = \operatorname{prox}_{\gamma f}\big(2x_n - y_n - \gamma \nabla h(x_n)\big)\\
y_{n+1} = y_n + \lambda_n (z_n - x_n).
\end{array}
\right.
\end{array} \tag{4.3}$$

Hence, we arrive at the following method.

Proposition 4.2. *Consider the setting of Problem 1.3. Assume that $\gamma \in\,]0, 2\beta[$ and that*

$$0 \in \operatorname{range}\big(\partial f + \partial \operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p} + \nabla h\big). \tag{4.4}$$

Set $\delta = 2 - \gamma/(2\beta)$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, \delta[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$, and let $y_0 \in \mathcal{H}$. Iterate

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} x_n = y_n - \sum_{k=1}^{p} \alpha_k L_k^*\big(L_k y_n - \operatorname{prox}_{\gamma g_k}(L_k y_n)\big)\\ z_n = \operatorname{prox}_{\gamma f}\big(2x_n - y_n - \gamma\nabla h(x_n)\big)\\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array}\right.\end{array} \tag{4.5}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution to Problem 1.3.

Proof. By Proposition 3.3(ii), the algorithms (4.3) and (4.5) are equivalent. Therefore, the assertion follows from [27, Theorem 2.1.1(b)] applied to (4.3). ☐

Heuristically, one can expect (4.5) to yield faster convergence than (4.2). Indeed, since the proximity operator of the aggregation $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}$ is computable in closed form via Proposition 3.3(ii), algorithm (4.3) processes only the three functions f, $\operatorname{pcm}_\gamma(g_k, L_k)_{1\leqslant k\leqslant p}$, and h. It also requires minimum variable storage. By contrast, the standard composite average function $\operatorname{cav}(g_k, L_k)_{1\leqslant k\leqslant p}$ in Problem 1.2 has no explicit proximity operator and its processing in algorithm (4.2) requires the splitting of the p functions $(g_k)_{1\leqslant k\leqslant p}$ and the p linear operator $(L_k)_{1\leqslant k\leqslant p}$ (the same holds true for all splitting methods involving standard composite averages [16]). Additionally, algorithm (4.2) requires the storage of significantly more variables than (4.5), in particular of the dual variables $(u_{k,n}^*)_{1\leqslant k\leqslant p}$.

§5. Numerical experiments

We illustrate the proposed proximal comixture model of Problem 1.3 through several experiments. The algorithms are executed with all initial vectors set to 0 and they use parameters η, σ, and $(\lambda_n)_{n\in\mathbb{N}}$ aiming at optimizing their performance. The stopping criterion to approximate the limit x_∞ of the sequence $(x_n)_{n\in\mathbb{N}}$ produced by an algorithm is $\|x_{n+1} - x_n\|/\|x_n\| < 10^{-6}$. The simulations are run using GNU Octave 8.4.0 on a laptop running Ubuntu 24.04.4. The code is available on the GitHub repository https://github.com/djcornejo/pcmm. The results of the these experiments support the fact that the proximal comixture model of Problem 1.3 leads to reliable solutions which can be obtained by efficient algorithms that are much less demanding in terms of memory requirements than those of Problem 1.2.

5.1. Experiment 1: Proximal total variation denoising

The purpose of this experiment is to illustrate the asymptotic property of Theorem 3.6(iii) on a simple denoising problem. Let $\overline{x} \in \mathbb{R}^N$ ($N = 256$) be the original 1-dimensional signal shown in Fig. 2(a) and let

$$z = \overline{x} + \frac{1}{2}w \tag{5.1}$$

be the noisy observation shown in Fig. 2(b), where $w \in \mathbb{R}^N$ is a realization of a normalized white Gaussian noise vector. We denote by $D\colon \mathbb{R}^N \to \mathbb{R}^N\colon (\xi_1, \ldots, \xi_N) \mapsto (\xi_2 - \xi_1, \ldots, \xi_N - \xi_{N-1}, \xi_1 - \xi_N)/2$ the normalized discrete gradient operator.

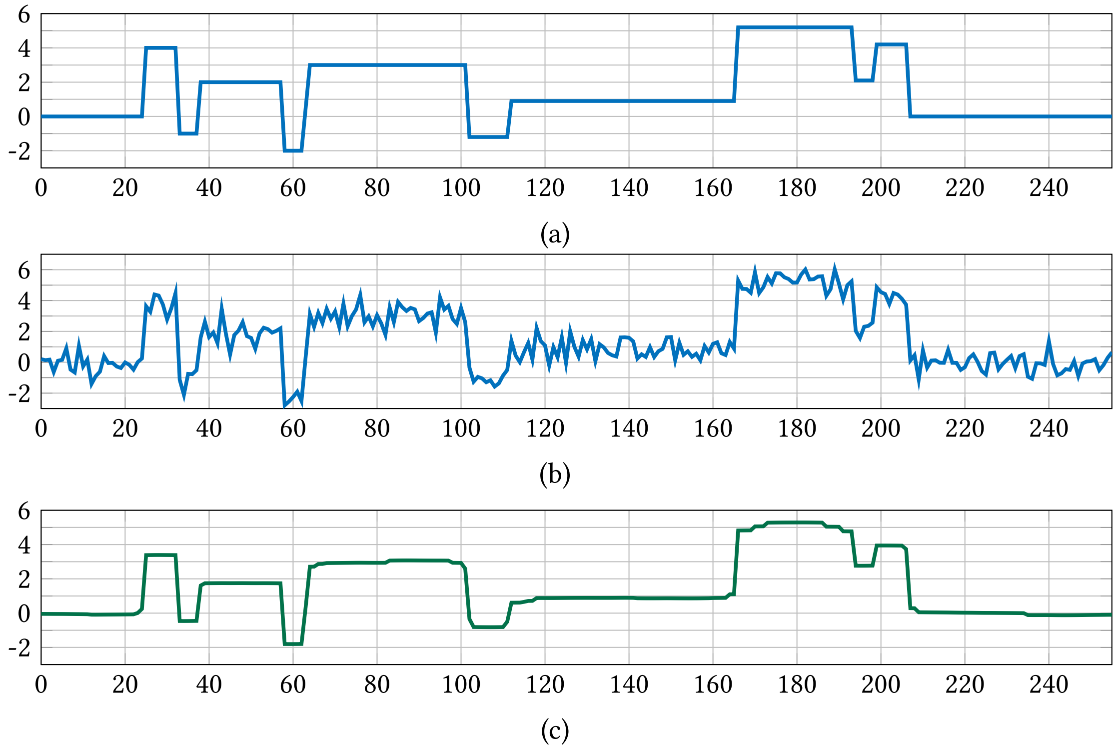


Figure 2: (a) Original signal $\overline{x}$. (b) Noisy observation z. (c) Solution to Problem 5.2 for $\gamma = 10^{-3}$ (the solution to Problem 5.1 is essentially identical).

Problem 5.1. Let $\rho = 3$ and let $\mathrm{tv} = \|\cdot\|_1 \circ D$ be the standard total variation loss. The task is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \ \mathrm{tv}(x) + \frac{1}{2\rho}\|x-z\|^2. \tag{5.2}$$

Problem 5.2. Let $\rho = 3$ and $\gamma \in \left]0,+\infty\right[$. Let us introduce the *proximal total variation* loss $\mathrm{ptv}_\gamma = D \overset{\gamma}{\blacklozenge} \|\cdot\|_1$. The task is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \ \mathrm{ptv}_\gamma(x) + \frac{1}{2\rho}\|x-z\|^2. \tag{5.3}$$

Problems 5.1 and 5.2 are particular instances of Problems 1.2 and 1.3, respectively, where $\mathcal{H} = \mathbb{R}^N$, $f = 0$, $h = \|\cdot - z\|^2/(2\rho)$, $p = 1$, $\alpha_1 = 1$, $g_1 = \|\cdot\|_1$, and $L_1 = D$. In view of Proposition 3.3(i) and (2.6), the unique solutions to Problems 5.1 and 5.2 are, respectively, $\mathrm{prox}_{\rho\,\mathrm{tv}}\, z$ and $\mathrm{prox}_{\rho\,\mathrm{ptv}_\gamma}\, z$. Therefore, Theorem 3.6(iii) asserts that the solution curve $(\mathrm{prox}_{\rho\,\mathrm{ptv}_\gamma}\, z)_{\gamma\in\left]0,+\infty\right[}$ in Problem 5.2 converges to the solution to Problem 5.1, to wit,

$$\mathrm{prox}_{\rho\,\mathrm{ptv}_\gamma}\, z \to \mathrm{prox}_{\rho\,\mathrm{tv}}\, z \quad \text{as} \quad \gamma \downarrow 0. \tag{5.4}$$

Since the proximity operator of tv is not known explicitly, algorithm (4.2) can be applied to obtain the solution to Problem 5.1. On the other hand, algorithm (4.5) can be used to obtain the solution to Problem 5.2. Our numerical experiments confirmed that, for $\gamma \leqslant 10^{-3}$, the solutions to Problems 5.1 and 5.2 were essentially identical, with a relative difference $\|\mathrm{prox}_{\rho\,\mathrm{tv}}\, z - \mathrm{prox}_{\rho\,\mathrm{ptv}_{10^{-3}}}\, z\|/\|\mathrm{prox}_{\rho\,\mathrm{tv}}\, z\| \approx 1.726 \times 10^{-3}$, in conformity with (5.4); see Fig. 2(c) for the denoised signal.

5.2. Experiment 2: Multiview image reconstruction

We address the problem of reconstructing the original image $\overline{x} \in C = [0, 255]^N$ ($N = 512^2$) of Fig. 3(a) from a partial observation of its possibly corrupted diffraction $r \approx \widehat{\overline{x}}$ over some frequency range R [41], where $\widehat{\overline{x}}$ denotes the two-dimensional discrete Fourier transform of $\overline{x}$. To exploit this imprecise information, we use the soft constraint distance penalty d_E, where

$$E = \left\{x \in \mathbb{R}^N \mid (\forall \nu \in R) \;\; \widehat{x}(\nu) = r(\nu)\right\}. \tag{5.5}$$

The set R contains the frequencies in $\{0, \ldots, 15\}^2$ as well as those resulting from the symmetry properties of the discrete Fourier transform. Further, two blurred noisy observations of $\overline{x}$ are available, namely (see Fig. 3(b)–(c))

$$z_1 = H_1\overline{x} + w_1 \quad \text{and} \quad z_2 = H_2\overline{x} + w_2, \tag{5.6}$$

where H_1 and H_2 model convolutional blurs with uniform rectangular kernels of sizes 14×18 and 20×5, respectively, while w_1 and w_2 represent zero-mean white Gaussian noise realizations of standard deviations of 2 and 3, respectively. The blurred image-to-noise ratios are 34.51 dB and 31.06 dB, respectively. Let

$$D \colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N \colon x \mapsto (D_1x, D_2x), \tag{5.7}$$

where D_1 and D_2 denote, respectively, the horizontal and vertical first order discrete difference operators, i.e.,

$$(D_1x)_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j}, & \text{if } j < 512;\\ 0, & \text{if } j = 512, \end{cases} \quad \text{and} \quad (D_2x)_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j}, & \text{if } i < 512;\\ 0, & \text{if } i = 512. \end{cases} \tag{5.8}$$

Let $\|\cdot\|_{1,2} \colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \colon (\xi_i, \eta_i)_{1\leqslant i\leqslant N} \mapsto \sum_{i=1}^N \sqrt{\xi_i^2 + \eta_i^2}$ and let

$$\mathfrak{h}_\rho \colon \mathbb{R} \to \mathbb{R} \colon \xi \mapsto \begin{cases} \rho|\xi| - \dfrac{\rho^2}{2}, & \text{if } \;\; |\xi| > \rho;\\ \dfrac{|\xi|^2}{2}, & \text{if } \;\; |\xi| \leqslant \rho \end{cases} \tag{5.9}$$

be the Huber function with parameter $\rho \in \left]0, +\infty\right[$. Our first formulation involves a standard composite average.

Problem 5.3. Let $\rho_1 = 3000$ and $\rho_2 = 4000$. The task is to

$$\underset{x \in C}{\text{minimize}} \;\; \left(\frac{1}{2}d_E(x) + \frac{1}{2}\|Dx\|_{1,2}\right) + \Big(\mathfrak{h}_{\rho_1}\big(\|H_1x - z_1\|\big) + \mathfrak{h}_{\rho_2}\big(\|H_2x - z_2\|\big)\Big). \tag{5.10}$$

The second formulation involves a proximal comixture.

Problem 5.4. Let $\rho_1 = 3000$, $\rho_2 = 4000$, and $\gamma \in \left]0, 1\right[$. The task is to

$$\underset{x \in C}{\text{minimize}} \;\; \Big(\mathrm{pcm}_\gamma\big(d_E, \mathrm{Id}; \sqrt{8}\|\cdot\|_{1,2}, D/\sqrt{8}\big)\Big)(x) + \Big(\mathfrak{h}_{\rho_1}\big(\|H_1x - z_1\|\big) + \mathfrak{h}_{\rho_2}\big(\|H_2x - z_2\|\big)\Big). \tag{5.11}$$

(a) (b) (c)

Figure 3: (a) Original image $\overline{x}$. (b) Degraded image z_1. (c) Degraded image z_2.

Problems 5.3 and 5.4 are particular instances of Problems 1.2 and 1.3, respectively, where $\mathcal{H} = \mathbb{R}^N$, $f = \iota_C$, $h = \mathfrak{h}_{\rho_1} \circ \|H_1 \cdot - z_1\| + \mathfrak{h}_{\rho_2} \circ \|H_2 \cdot - z_2\|$, $\beta = 1/2$, $p = 2$, $\alpha_1 = \alpha_2 = 1/2$, $\mathcal{G}_1 = \mathbb{R}^N$, $g_1 = d_E$, $L_1 = \mathrm{Id}$, $\mathcal{G}_2 = \mathbb{R}^N \times \mathbb{R}^N$, $g_2 = \sqrt{8}\|\cdot\|_{1,2}$, and $L_2 = D/\sqrt{8}$. In this case, $\|L_1\|^2 = \|L_2\|^2 = 1$ and

$$\operatorname{prox}_{\gamma f} \colon \mathbb{R}^N \to \mathbb{R}^N \colon (\xi_i)_{1\leqslant i\leqslant N} \mapsto \big(\min\big\{\max\{\xi_i, 0\}, 255\big\}\big)_{1\leqslant i\leqslant N}. \tag{5.12}$$

Moreover, [5, Example 24.28] yields

$$\operatorname{prox}_{\gamma g_1} \colon \mathbb{R}^N \to \mathbb{R}^N \colon x \mapsto \begin{cases} x + \dfrac{\gamma}{d_E(x)}(\operatorname{proj}_E x - x), & \text{if } d_E(x) > \gamma;\\ \operatorname{proj}_E x, & \text{if } d_E(x) \leqslant \gamma, \end{cases} \tag{5.13}$$

where

$$\operatorname{proj}_E x = \mathsf{IDFT}(\widetilde{x}), \quad \text{with} \quad \widetilde{x}(\nu) = \begin{cases} r(\nu), & \text{if } \nu \in R;\\ \widehat{x}(\nu), & \text{otherwise.} \end{cases} \tag{5.14}$$

Further, [5, Proposition 24.11] yields

$$\operatorname{prox}_{\gamma g_2} \colon \mathbb{R}^N\times\mathbb{R}^N \to \mathbb{R}^N\times\mathbb{R}^N \colon (\xi_i, \eta_i)_{1\leqslant i\leqslant N} \mapsto \big(\varrho_i\xi_i, \varrho_i\eta_i\big)_{1\leqslant i\leqslant N}, \text{ where } \varrho_i = 1 - \frac{\sqrt{8}\gamma}{\max\big\{\sqrt{8}\gamma, \|(\xi_i, \eta_i)\|\big\}}. \tag{5.15}$$

At last, [19, Example 2.3] establishes that

$$\nabla h \colon \mathbb{R}^N \to \mathbb{R}^N \colon x \mapsto \frac{\rho_1 H_1^*(H_1 x - z_1)}{\max\{\rho_1, \|H_1 x - z_1\|\}} + \frac{\rho_2 H_2^*(H_2 x - z_2)}{\max\{\rho_2, \|H_2 x - z_2\|\}}. \tag{5.16}$$

We construct the solution to Problem 5.3 shown in Fig. 4(a) via Proposition 4.1, where $\eta = 1.99\beta$ and $\sigma = 0.99/\eta$. To apply Proposition 4.2 to Problem 5.4, let us verify that condition (4.4) is satisfied. By Proposition 3.3(iii)(c) and [5, Corollary 16.48(iii)],

$$\partial\Big(f + \operatorname{pcm}_\gamma\big(d_E, \mathrm{Id}; \sqrt{8}\|\cdot\|_{1,2}, D/\sqrt{8}\big) + h\Big) = \partial f + \partial \operatorname{pcm}_\gamma\big(d_E, \mathrm{Id}; \sqrt{8}\|\cdot\|_{1,2}, D/\sqrt{8}\big) + \nabla h, \tag{5.17}$$

which is a maximally monotone operator with domain C [5, Theorem 20.25]. Hence, (4.4) follows from [5, Corollary 21.25]. The image reconstructed by (4.5) for Problem 5.4 with $\gamma = 0.1$ and $\lambda_n \equiv$

(a) Problem 5.3. (b) Problem 5.4 ($\gamma = 0.1$). (c) Problem 5.4 ($\gamma = 0.99$).

Figure 4: Images reconstructed by Problems 5.3 and 5.4.

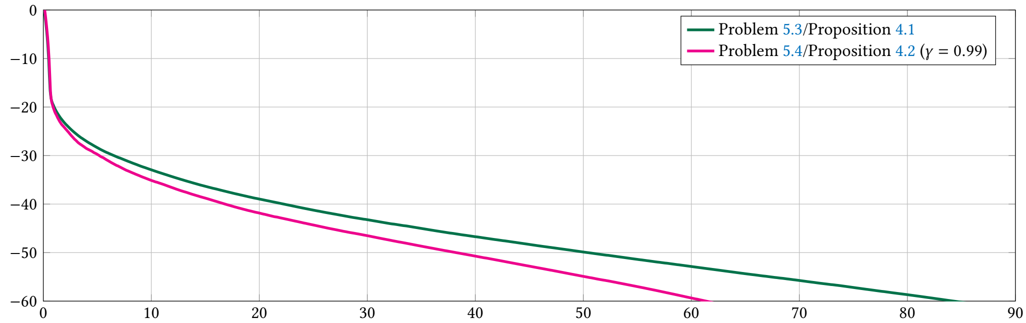


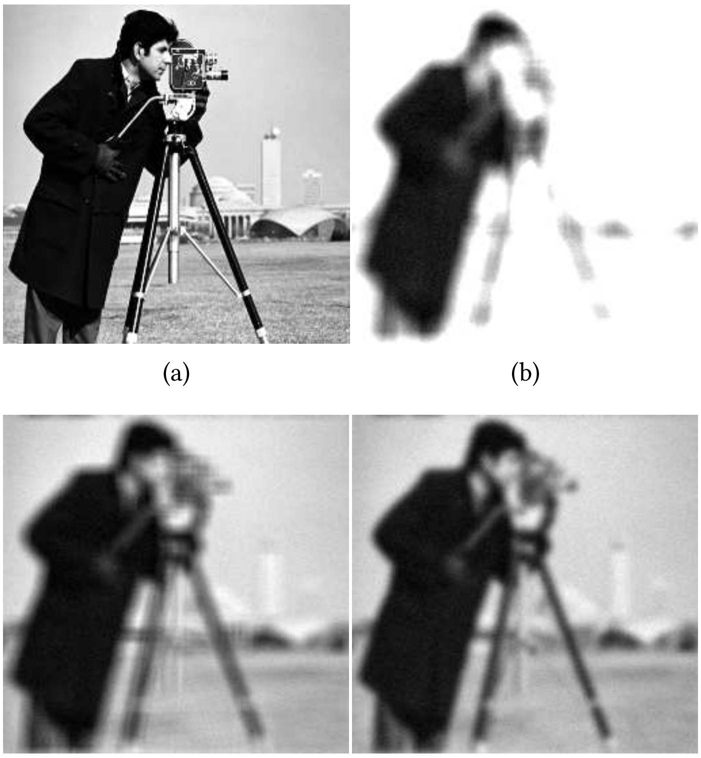
Figure 5: Normalized error $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus time (s).

$1.89 < 1.90 = 2-\gamma/(2\beta)$ is shown in Fig. 4(b). As predicted by Theorem 3.6(iv)(b), since γ is small, this solution is similar to that produced by Problem 5.3 in Fig. 4(a). The solution produced by Problem 5.4 for $\gamma = 0.99$ with $\lambda_n \equiv 1 < 1.01 = 2 - \gamma/(2\beta)$ in (4.5) is shown in Fig. 4(c) to yield a slightly sharper reconstruction. Finally, Fig. 5 illustrates the faster convergence of algorithm (4.5) for the proximal comixture model (5.11).

5.3. Experiment 3: Image reconstruction from phase

We address a phase recovery problem considered in [19]. The goal is to recover the original image $\overline{x} \in C = [0, 255]^N$ ($N = 256^2$) shown in Fig. 6(a) from an imprecise observation of its Fourier phase $\theta \approx \angle\widehat{\overline{x}}$ [35]. The problem is modeled as a convex feasibility problem with the following constraint sets.

- Phase: $C_1 = \big\{x \in \mathbb{R}^N \mid \angle\,\widehat{x} = \theta\big\}$.
- Proximity to the reference image r of Fig. 6(b): $C_2 = \big\{x \in \mathbb{R}^N \mid \|x - r\|_2 \leqslant \xi\big\}$. The image r is a blurred and noise-corrupted version of $\overline{x}$, which is further degraded by saturation (the pixel values beyond 130 are clipped to 130).
- Upper bound on the norm of the gradient: $Dx/\sqrt{8} \in C_3$, where D is defined as in (5.7) and $C_3 = \big\{y \in \mathbb{R}^N \times \mathbb{R}^N \mid \|y\|_2 \leqslant \rho\big\}$.
- A blurred observation $z_1 = H_1\overline{x} + w_1$ is available (see Fig. 6(c)), where H_1 models a convolutional blur with a uniform rectangular kernel of size 11×13 and w_1 is a realization of a bounded random



(a) (b) (c) (d)

Figure 6: (a) Original image $\overline{x}$. (b) Reference image r. (c) Degraded image z_1. (d) Degraded image z_2.

noise vector with components in $[-\eta, \eta]$, where $\eta = 7$. This yields a blurred image-to-noise ratio of 31.90 dB. A candidate solution $x \in \mathbb{R}^N$ should be such that, for every $k \in \{1, \ldots, N\}$, the kth component of $H_1 x - z_1$ lies in $[-\eta, \eta]$, that is

$$\langle H_1 x \mid e_k \rangle \in C_{k+3} = \big[\langle z_1 \mid e_k \rangle - \eta, \langle z_1 \mid e_k \rangle + \eta\big], \tag{5.18}$$

where $(e_k)_{1 \leqslant k \leqslant N}$ is the standard orthonormal basis of $\mathbb{R}^N$.

- A second blurred observation of $\overline{x}$ is available, namely (see Fig. 6(d)) $z_2 = H_2 \overline{x} + w_2$, where H_2 models a 17×17 convolutional blur obtained by truncating the support of a Gaussian kernel with standard deviation 3, and w_2 is a realization of an unknown white Gaussian noise. The blurred image-to-noise ratio is 30.12 dB. This information is penalized by the function $h \colon x \mapsto \|H_2 x - z_2\|^2/2$.

Because of inaccuracies in the values θ, ξ, and ρ, the convex feasibility problem arising from the above constraints might be inconsistent and we relax it using the Berhu function given by

$$\mathfrak{b} \colon \mathbb{R} \to \mathbb{R} \colon \xi \mapsto \begin{cases} \dfrac{\xi^2 + 1}{2}, & \text{if } |\xi| > 1; \\ |\xi|, & \text{if } |\xi| \leqslant 1. \end{cases} \tag{5.19}$$

In addition, we set $L_1 = L_2 = \mathrm{Id}_{\mathbb{R}^N}$, $L_3 = D/\sqrt{8}$, and $\alpha_1 = \alpha_2 = \alpha_3 = 1/4$ while, for every $k \in \{1, \ldots, N\}$, $L_{k+3} \colon \mathbb{R}^N \to \mathbb{R} \colon x \mapsto \langle H_1 x \mid e_k \rangle$ and $\alpha_{k+3} = 1/(4N)$.

Our first formulation employs a standard composite average.

(a) Problem 5.5. (b) Problem 5.6 ($\gamma = 0.1$). (c) Problem 5.6 ($\gamma = 1.99$).

Figure 7: Images reconstructed by Problems 5.5 and 5.6.

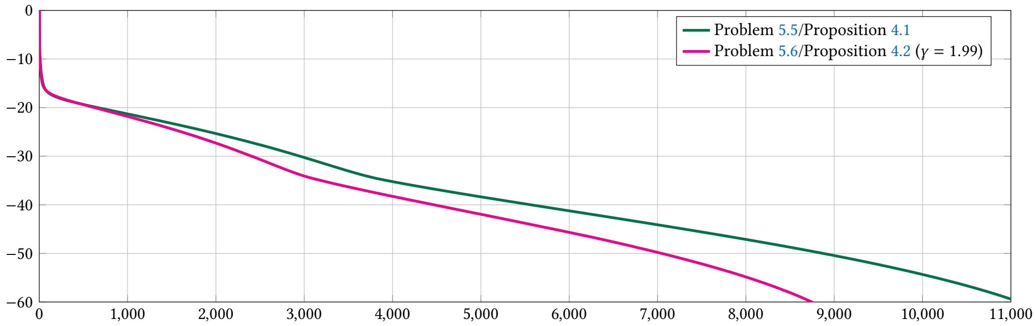


Figure 8: Normalized error $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus time (s).

Problem 5.5. The task is to

$$\underset{x\in C}{\text{minimize}} \quad \sum_{k=1}^{N+3} \alpha_k \mathfrak{b}\big(d_{C_k}(L_k x)\big) + \frac{1}{2}\|H_2 x - z_2\|^2. \tag{5.20}$$

Our second formulation employs a proximal comixture.

Problem 5.6. Let $\gamma \in\,]0,2[$. The task is to

$$\underset{x\in C}{\text{minimize}} \quad \Big(\mathrm{pcm}_\gamma\big(\mathfrak{b}\circ d_{C_k}, L_k\big)_{1\leqslant k\leqslant N+3}\Big)(x) + \frac{1}{2}\|H_2 x - z_2\|^2. \tag{5.21}$$

Problems 5.5 and 5.6 are particular instances of Problems 1.2 and 1.3, respectively, where $\mathcal{H} = \mathbb{R}^N$, $f = \iota_C$, $h = \|H_2\cdot - z_2\|^2/2$, $\beta = 1$, $p = N+3$, and, for every $k \in \{1,\ldots,p\}$, $g_k = \mathfrak{b}\circ d_{C_k}$. We have $(\forall k \in \{1,\ldots,p\})\ \|L_k\|^2 \leqslant 1$. Further, $\mathrm{prox}_{\gamma f}$ is given as in (5.12), [5, Proposition 24.27] yields

$$(\forall k \in \{1,2\}) \quad \mathrm{prox}_{\gamma g_k} \colon \mathbb{R}^N \to \mathbb{R}^N \colon x \mapsto \begin{cases} \mathrm{proj}_{C_k} x + \dfrac{1}{1+\gamma}\big(x - \mathrm{proj}_{C_k} x\big), & \text{if } d_{C_k}(x) > 1+\gamma;\\ x + \dfrac{\gamma}{d_{C_k}(x)}\big(\mathrm{proj}_{C_k} x - x\big), & \text{if } \gamma < d_{C_k}(x) \leqslant 1+\gamma;\\ \mathrm{proj}_{C_k} x, & \text{if } d_{C_k}(x) \leqslant \gamma, \end{cases} \tag{5.22}$$

as well as

$$\operatorname{prox}_{\gamma g_3}\colon \mathbb{R}^N\times\mathbb{R}^N \to \mathbb{R}^N\times\mathbb{R}^N\colon y \mapsto \begin{cases} \operatorname{proj}_{C_3} y + \dfrac{1}{1+\gamma}(y - \operatorname{proj}_{C_3} y), & \text{if } d_{C_3}(y) > 1+\gamma;\\ y + \dfrac{\gamma}{d_{C_3}(y)}(\operatorname{proj}_{C_3} y - y), & \text{if } \gamma < d_{C_3}(y) \leqslant 1+\gamma;\\ \operatorname{proj}_{C_3} y, & \text{if } d_{C_3}(y) \leqslant \gamma \end{cases} \tag{5.23}$$

and

$$(\forall k \in \{4,\ldots,N{+}3\})\;\; \operatorname{prox}_{\gamma g_k}\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} \operatorname{proj}_{C_k} \xi + \dfrac{1}{1+\gamma}(\xi - \operatorname{proj}_{C_k} \xi), & \text{if } d_{C_k}(\xi) > 1+\gamma;\\ \xi + \dfrac{\gamma}{d_{C_k}(\xi)}(\operatorname{proj}_{C_k} \xi - \xi), & \text{if } \gamma < d_{C_k}(\xi) \leqslant 1+\gamma;\\ \operatorname{proj}_{C_k} \xi, & \text{if } d_{C_k}(\xi) \leqslant \gamma, \end{cases} \tag{5.24}$$

where $\operatorname{proj}_{C_1}$ is given by [35]

$$\operatorname{proj}_{C_1} x = \mathsf{IDFT}\Big(|\mathsf{DFT}x| \max\{\cos\big(\angle(\mathsf{DFT}x) - \theta\big), 0\} \exp(\imath\theta)\Big), \tag{5.25}$$

and, for every $k \in \{2,\ldots,N+3\}$, $\operatorname{proj}_{C_k}$ follows from [5, Example 3.18 and Proposition 3.19]. Finally, $\nabla h = H_2^* \circ (H_2 \cdot - z_2)$.

We apply Proposition 4.1 to Problem 5.5 with $\eta = 1.99\beta$ and $\sigma = 0.99/\eta$, which produces the reconstructed image shown in Fig. 7(a). Following the same argument used in Problem 5.4, we note that condition (4.4) is satisfied and we apply Proposition 4.2 to Problem 5.6. The image reconstructed by (4.5) for $\gamma = 0.1$ and $\lambda_n \equiv 1.94 < 1.95 = 2 - \gamma/(2\beta)$ is shown in Fig. 7(b). This solution is similar to that of Fig. 7(a), which is consistent with Theorem 3.6(iv)(b) since γ is small. For Problem 5.6 with $\gamma = 1.99$, algorithm (4.5) with $\lambda_n \equiv 1 < 1.005 = 2-\gamma/(2\beta)$ yields the somewhat sharper reconstruction shown in Fig. 7(c). Fig. 8 depicts the faster convergence of algorithm (4.5) for the proximal comixture model (5.21).

5.4. Experiment 4: Linear regression

This experiment focuses on a linear regression model with an overlapping group structure on the inputs [12, 49]. We consider $p = 100$ groups of indices $(I_k)_{1\leqslant k\leqslant p}$ in $\{1,\ldots,N\}$ ($N = 90p + 10 = 9010$) of length 100 such that two consecutive groups overlap by 10 variables, i.e.,

$$I_1 = \{1,\ldots,100\}, I_2 = \{91,\ldots,190\},\ldots, I_p = \{N-99,\ldots,N\}. \tag{5.26}$$

We use $M = 10000$ samples. The entries of the input matrix $A \in \mathbb{R}^{M\times N}$ are i.i.d. samples from a $\mathcal{N}(0,1)$ distribution. The entries of the true regression coefficients $\overline{x} \in \mathbb{R}^N$ are also i.i.d. samples from a $\mathcal{N}(0,1)$ distribution, and the output data is generated by the noisy linear model $z = A\overline{x} + w$, where $w \in \mathbb{R}^M$ has entries that are i.i.d. samples from a $\mathcal{N}(0,1)$ distribution. For every $k \in \{1,\ldots,p\}$, we set $L_k\colon \mathbb{R}^N \to \mathbb{R}^{100}\colon x = (\xi_i)_{1\leqslant i\leqslant N} \mapsto (\xi_i)_{i\in I_k}$.

As before, we consider formulations based on a standard composite average and the proximal comixture.

Problem 5.7. The task is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\;\; \frac{1}{p}\|x\|_1 + \frac{1}{p}\sum_{k=1}^{p} \|L_k x\| + \frac{1}{2p^2}\|Ax - z\|^2. \tag{5.27}$$

Problem 5.8. Let $\gamma \in \left]0, 2p^2/\|A\|^2\right[$. The task is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\;\; \frac{1}{p}\|x\|_1 + \Big(\mathrm{pcm}_\gamma\big(\|\cdot\|, L_k\big)_{1\leqslant k\leqslant p}\Big)(x) + \frac{1}{2p^2}\|Ax-z\|^2. \tag{5.28}$$

Problems 5.7 and 5.8 are particular instances of Problems 1.2 and 1.3, respectively, where $\mathcal{H} = \mathbb{R}^N$, $f = \|\cdot\|_1/p$, $h = \|A\cdot -z\|^2/(2p^2)$, $\beta = p^2/\|A\|^2$, and, for every $k \in \{1,\ldots,p\}$, $\alpha_k = 1/p$, $\mathcal{G}_k = \mathbb{R}^{100}$, $g_k = \|\cdot\|$, and $\|L_k\| = 1$. Additionally,

$$\mathrm{prox}_{\gamma f}\colon \mathbb{R}^N \to \mathbb{R}^N\colon (\xi_i)_{1\leqslant i\leqslant N} \mapsto \big(\mathrm{sign}\,(\xi_i)\max\{|\xi_i| - \gamma/p, 0\}\big)_{1\leqslant i\leqslant N}. \tag{5.29}$$

Further, [5, Example 24.20] yields

$$(\forall k\in\{1,\ldots,p\})\;\; \mathrm{prox}_{\gamma g_k}\colon \mathbb{R}^{100} \to \mathbb{R}^{100}\colon y \mapsto \left(1 - \frac{\gamma}{\max\{\|y\|,\gamma\}}\right)y, \tag{5.30}$$

while $\nabla h = (1/p^2)A^* \circ (A\cdot -z)$. Next, let us verify that condition (4.4) is satisfied. Note that, by Proposition 3.5, $\mathrm{pcm}_\gamma(\|\cdot\|, L_k)_{1\leqslant k\leqslant p} \geqslant 0$. Thus, the objective function of Problem 5.8 is coercive and the existence of minimizers is guaranteed by [5, Proposition 11.15(i)]. Therefore, Fermat's rule [5, Theorem 16.3] and [5, Corollary 16.48(iii)] guarantee that condition (4.4) holds. The solutions x_∞ to Problem 5.7 and to Problem 5.8 for $\gamma = 1.99\beta$ are very close, with a relative difference $\|x_{\infty,(5.27)} - x_{\infty,(5.28)}\|/\|x_{\infty,(5.27)}\| \approx 5.045\times 10^{-6}$. In addition, $\|x_\infty - \overline{x}\|/\|\overline{x}\| \approx 0.153$. For Problem 5.7, we use algorithm (4.2) with $\eta = 1.99\beta$ and $\sigma = 0.99/\eta$, while for Problem 5.8 for $\gamma = 1.99\beta$, we use algorithm (4.5) with $\lambda_n \equiv 1 < 1.005 = 2 - \gamma/(2\beta)$. The convergence pattern of algorithm (4.5) for the proximal comixture model (5.28) is shown in Fig. 9.

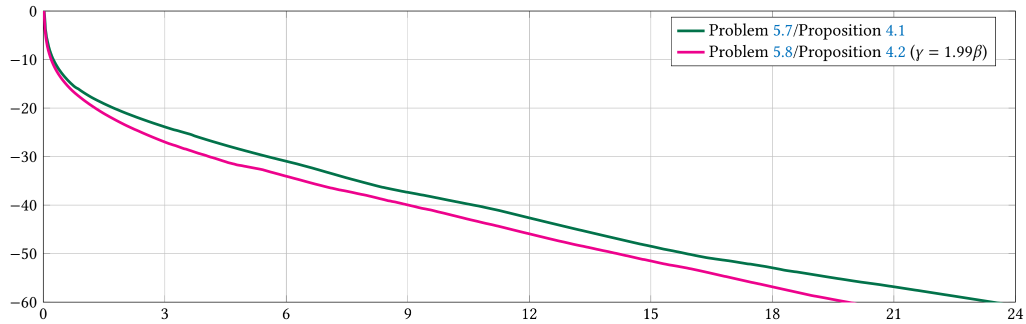


Figure 9: Normalized error $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus time (s).

References

[1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Prentice-Hall, Englewood Cliffs, NJ, 1977.

[2] A. Ang, W. Bin Hamed, and H. De Sterck, Sum-of-norms regularized nonnegative matrix factorization, *Neural Comput.*, vol. 38, pp. 228–255, 2026.

[3] H. Attouch, *Variational Convergence for Functions and Operators.* Pitman, Boston, MA, 1984.

[4] H. Attouch and D. Azé, Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry–Lions method, *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, vol. 10, pp. 289–312, 1993.

[5] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, New York, 2017.

[6] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, The proximal average: Basic theory, *SIAM J. Optim.*, vol. 19, pp. 766–785, 2008.

[7] P. Bernard, Lasry–Lions regularization and lemma of Ilmanen, *Rend. Semin. Mat. Univ. Padova*, vol. 124, pp. 221–229, 2010.

[8] M. N. Bùi and P. L. Combettes, Integral resolvent and proximal mixtures, *J. Optim. Theory Appl.*, vol. 203, pp. 2328–2353, 2024.

[9] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld, The multiple-sets split feasibility problem and its applications for inverse problems, *Inverse Problems*, vol. 21, pp. 2071–2084, 2005.

[10] Y. Censor and M. Zaknoon, Algorithms and convergence results of projection methods for inconsistent feasibility problems: A review, *Pure Appl. Funct. Anal.*, vol. 3, pp. 565–586, 2018.

[11] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numer.*, vol. 25, pp. 161–319, 2016.

[12] X. Chen, Q. Lin, S. Kim, J. G. Carbonell, and E. P. Xing, Smoothing proximal gradient method for general structured sparse regression, *Ann. Appl. Stat.*, vol. 6, pp. 719–752, 2012.

[13] Y.-M. Cheung and J. Lou, Proximal average approximated incremental gradient descent for composite penalty regularized empirical risk minimization, *Mach. Learn.*, vol. 106, pp. 595–622, 2017.

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

[15] P. L. Combettes, Resolvent and proximal compositions, *Set-Valued Var. Anal.*, vol. 31, art. 22, 29 pp., 2023.

[16] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[17] P. L. Combettes and D. J. Cornejo, Signal recovery with proximal comixtures, *Proc. Europ. Signal Process. Conf.*, pp. 2637–2641. Lyon, France, August 26–30, 2024.

[18] P. L. Combettes and D. J. Cornejo, Variational analysis of proximal compositions and integral proximal mixtures. *Evol. Equ. Control Theory*, vol. 17, pp. 106–139, 2026.

[19] P. L. Combettes and L. E. Glaudin, Proximal activation of smooth functions in splitting algorithms for convex image recovery, *SIAM J. Imaging Sci.*, vol. 12, pp. 1905–1935, 2019.

[20] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, New York, 2011.

[21] P. L. Combettes and J.-C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite Lipschitzian, and parallel-sum type monotone operators, *Set-Valued Var. Anal.*, vol. 20, pp. 307–330, 2012.

[22] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

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

[24] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *J. Optim. Theory Appl.*, vol. 158, pp. 460–479, 2013.

[25] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists, *SIAM Rev.*, vol. 65, pp. 375–435, 2023.

[26] L. Condat, G. Malinovsky, and P. Richtárik, Distributed proximal splitting algorithms with rates and acceleration, *Front. Signal Process.*, vol. 1, art. 776825, 2022.

[27] D. Davis and W. Yin, A three-operator splitting scheme and its optimization applications, *Set-Valued Var. Anal.*, vol. 25, pp. 829–858, 2017.

[28] R. Glowinski, S. J. Osher, and W. Yin (eds.), *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer, New York, 2016.

[29] J.-B. Hiriart-Urruty, A general formula on the conjugate of the difference of functions, *Canad. Math. Bull.*, vol. 29, pp. 482–485, 1986.

[30] J.-B. Hiriart-Urruty, The deconvolution operation in convex analysis: An introduction, *Cybernet. Systems Anal.*, vol. 30, pp. 97–104, 1994.

[31] Z. Hu, K. Shaloudegi, G. Zhang, and Y. Yu, Federated learning meets multi-objective optimization, *IEEE Trans. Network Sci. Eng.*, vol. 9, pp. 2039–2051, 2022.

[32] A. B. Juditsky and A. S. Nemirovski, *Statistical Inference via Convex Optimization*. Princeton University Press, Princeton, NJ, 2020.

[33] J. M. Lasry and P. L. Lions, A remark on regularization in Hilbert spaces, *Israel J. Math.*, vol. 55, pp. 257–266, 1986.

[34] A. M. Legendre, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Firmin Didot, Paris, 1805.

[35] A. Levi and H. Stark, Signal restoration from phase by projection onto convex sets, *J. Opt. Soc. Am.*, vol. 73, pp. 810–822, 1983.

[36] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, Federated optimization in heterogeneous networks, *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, 2020.

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

[38] P. Pérez-Aros and E. Vilches, Moreau envelope of supremum functions with applications to infinite and stochastic programming, *SIAM J. Optim.*, vol. 31, pp. 1635–1657, 2021.

[39] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, *J. Assoc. Comput. Mach.*, vol. 9, pp. 84–97, 1962.

[40] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, 3rd printing. Springer, New York, 2009.

[41] M. I. Sezan and H. Stark, Image restoration by convex projections in the presence of noise, *Appl. Opt.*, vol. 22, pp. 2781–2789, 1983.

[42] L. Shen, W. Liu, J. Huang, Y.-G. Jiang, and S. Ma, Adaptive proximal average approximation for composite convex minimization, *Proc. AAAI Conf. Artif. Intell.*, vol. 31, pp. 2513–2519, 2017.

[43] T. Strömberg, On regularization in Banach spaces, *Ark. Mat.*, vol. 34, pp. 383–406, 1996.

[44] Y. Sun, Z. Wu, X. Xu, B. Wohlberg, and U. S. Kamilov, Scalable plug-and-play ADMM with convergence guarantees, *IEEE Trans. Comput. Imaging*, vol. 7, pp. 849–863, 2021.

[45] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 2nd ed. Elsevier, Amsterdam, 2020.

[46] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[47] M. Yan, A new primal-dual algorithm for minimizing the sum of three functions with a linear operator, *J. Sci. Comput.*, vol. 76, pp. 1698–1717, 2018.

[48] Y.-L. Yu, Better approximation and faster algorithm using the proximal average, *Proc. Conf. Adv. Neural Inform. Process. Syst.*, pp. 458–466, 2013.

[49] L. Yuan, J. Liu, and J. Ye, Efficient methods for overlapping group lasso, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, pp. 2104–2116, 2013.