Signal Recovery with Proximal Comixtures

Patrick L. Combettes
Department of Mathematics
North Carolina State University
Raleigh, NC 27695-8205, USA

Diego J. Cornejo
Department of Mathematics
North Carolina State University
Raleigh, NC 27695-8205, USA

***Abstract*—In variational signal processing and machine learning problems, loss functions and linear operators are typically aggregated as an average of composite terms. We propose an alternative formulation using proximal comixtures, an operation that combines functions and linear operators in such a way that the proximity operator of the resulting function is computable explicitly. The benefits of comixture formulations are illustrated through image recovery and machine learning applications.**

***Index Terms*—proximal comixture, convex optimization, signal recovery.**

I. Introduction

Various data analysis problems in signal processing and machine learning can be condensed into the minimization of an aggregation of loss functions that model individually desired properties of the ideal solution in a Hilbert space $\mathcal{H}$. These properties typically result from prior knowledge and the observation of data. To be more specific, let us state our assumptions on the variational models to be discussed (see Section II for notation).

Assumption 1 $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \,|\, \cdot \rangle$ and associated norm $\|\cdot\|$, $f \in \Gamma_0(\mathcal{H})$, 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 $L_k\colon \mathcal{H} \to \mathcal{G}_k$ is a bounded linear operator such that (without loss of generality) $\|L_k\| \leqslant 1$. Further, the coefficients $(\alpha_k)_{1\leqslant k\leqslant p} \in \left]0,1\right]^p$ satisfy $\sum_{k=1}^p \alpha_k = 1$.

The most prevalent optimization framework used in data analysis problems is the following, in which the objective is to minimize the sum of a function f and p composite functions aggregated via a standard averaging operation (see [6], [11], [13] and the references therein).

Problem 2 Under Assumption 1, the task is to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} \alpha_k g_k(L_k x). \tag{1}$$

While simple from a modeling viewpoint, the above averaging process brings some challenge on the numerical side. Indeed, since the proximity operator of the composite average has no closed form expression, solving Problem 2 requires splitting $p+1$ terms, which typically leads to algorithms that are slower and necessitate more memory storage than those that would split less terms. In addition, the aggregation model of Problem 2 may not be robust to perturbations. For instance, let us consider the special case when $f = 0$ and each g_k is the indicator function of a nonempty closed convex set $D_k \subset \mathcal{G}_k$. This reduces (1) to the convex feasibility problem

$$\text{find} \;\; x \in \mathcal{H} \;\; \text{such that} \;\; (\forall k \in \{1,\ldots,p\}) \;\; L_k x \in D_k. \tag{2}$$

If the sets $(D_k)_{1\leqslant k\leqslant p}$ or the operators $(L_k)_{1\leqslant k\leqslant p}$ are not specified exactly, no solution may exist [4], [8].

The objective of the present paper is to propose the use of a new aggregation process, called the *proximal comixture*, to combine the functions $(g_k)_{1\leqslant k\leqslant p}$ and the linear operators $(L_k)_{1\leqslant k\leqslant p}$. This operation, introduced in [9], further studied in [3], and applied for the first time in the present paper, can be viewed as a generalization of the proximal average [1], [2], which corresponds to the special case in which, for every $k \in \{1,\ldots,p\}$, $\mathcal{G}_k = \mathcal{H}$ and $L_k = \mathrm{Id}$. In this specific context, the benefits of using proximal averages in lieu of standard averages has been documented in several studies, e.g., [14], [17], [19], [21]. We shall show that, more generally, solving minimization problems involving proximal comixtures instead of the composite averages of Problem 2 may yield notable modeling and computational advantages. For instance, as discussed above, the computation of the proximity operator of the standard average in (1) is not tractable and solving Problem 2 requires sophisticated splitting techniques. By contrast, the proximity operator of the comixture will be shown to be computable explicitly in terms of the individual proximity operators of the functions $(g_k)_{1\leqslant k\leqslant p}$. As a result, Problem 2 can be solved by splitting only two terms, namely f and the proximal comixture.

The remainder of the paper is organized as follows. Section II provides the necessary mathematical background and notation. Section III is devoted to proximal comixtures and their main properties. The proximal comixture minimization problem is introduced in Section IV. Finally, numerical experiments are presented in Section V.

II. Mathematical background and notation

Our notation follows [1], where one will find the necessary background. We denote by $\Gamma_0(\mathcal{H})$ the class of lower

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

semicontinuous convex functions $f\colon \mathcal{H} \to]-\infty,+\infty]$ which are proper, i.e., $\operatorname{dom} f = \big\{x \in \mathcal{H} \mid f(x) < +\infty\big\} \neq \varnothing$. Let $C \subset \mathcal{H}$. Then ι_C denotes the indicator function of C and d_C the distance function to the set C. Let $f \in \Gamma_0(\mathcal{H})$. The conjugate of f is the function $f^* \in \Gamma_0(\mathcal{H})$ defined by

$$f^*\colon \mathcal{H} \to]-\infty,+\infty] : u \mapsto \sup_{x\in\mathcal{H}} \big(\langle x \,|\, u\rangle - f(x)\big) \tag{3}$$

and the subdifferential of f at $x \in \mathcal{H}$ is the set

$$\partial f(x) = \big\{u \in \mathcal{H} \,|\, (\forall z \in \mathcal{H})\ \langle z - x \,|\, u\rangle + f(x) \leqslant f(z)\big\}. \tag{4}$$

Now let $\mathscr{Q}_{\mathcal{H}} = \|\cdot\|^2/2$ be the normalized quadratic kernel of $\mathcal{H}$. The Moreau envelope of f is

$$f \,\square\, \mathscr{Q}_{\mathcal{H}}\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \inf_{z\in\mathcal{H}} \big(f(z) + \mathscr{Q}_{\mathcal{H}}(x - z)\big) \tag{5}$$

and the proximity operator of f is

$$\operatorname{prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname*{argmin}_{z\in\mathcal{H}} \big(f(z) + \mathscr{Q}_{\mathcal{H}}(x - z)\big). \tag{6}$$

The *Huber function with parameter* $\rho \in]0,+\infty[$ is

$$\mathfrak{h}_\rho\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} \rho|\xi| - \dfrac{\rho^2}{2}, & \text{if } |\xi| > \rho; \\ \dfrac{|\xi|^2}{2}, & \text{if } |\xi| \leqslant \rho. \end{cases} \tag{7}$$

III. Proximal comixtures

Proximal comixtures were introduced in [9] and further investigated in [3] as a new operation that combines functions and linear operators.

Definition 3 Suppose that Assumption 1 is in force. The *proximal comixture* of $(g_k)_{1\leqslant k\leqslant p}$ and $(L_k)_{1\leqslant k\leqslant p}$ is

$$\mathop{\blacklozenge}\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p} = \left(\left(\sum_{k=1}^{p} \alpha_k (g_k \,\square\, \mathscr{Q}_{\mathcal{G}_k}) \circ L_k\right)^* - \mathscr{Q}_{\mathcal{H}}\right)^*.$$

Let us recall from [9] some key properties of proximal comixtures, in particular the fact that their proximity operator can be computed explicitly.

Proposition 4 *Suppose that Assumption 1 is in force and set* $h = \mathop{\blacklozenge}\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p}$. *Then the following hold:*

(i) $h \in \Gamma_0(\mathcal{H})$.
(ii) $\operatorname{prox}_h = \operatorname{Id} - \sum_{k=1}^{p} \alpha_k L_k^* \circ (\operatorname{Id} - \operatorname{prox}_{g_k}) \circ L_k$.
(iii) $\operatorname{Argmin} h = \operatorname{Argmin} \sum_{k=1}^{p} \alpha_k (g_k \,\square\, \mathscr{Q}_{\mathcal{G}_k}) \circ L_k$.

Let us provide a few illustrations of Definition 3, starting with the extreme case when it happens to coincide with the standard composite average.

Example 5 Suppose that Assumption 1 is in force and let $\boldsymbol{\mathcal{G}}$ be the standard product vector space $\mathcal{G}_1 \times \cdots \times \mathcal{G}_k$, with generic element $\boldsymbol{y} = (y_k)_{1\leqslant k\leqslant p}$, and equipped with the scalar product $(\boldsymbol{y}, \boldsymbol{v}) \mapsto \sum_{k=1}^{p} \alpha_k \langle y_k \,|\, v_k\rangle$. Suppose that $\boldsymbol{L}\colon \mathcal{H} \to \boldsymbol{\mathcal{G}}\colon x \mapsto (L_k x)_{1\leqslant k\leqslant p}$ satisfies $\boldsymbol{L} \circ \boldsymbol{L}^* = \operatorname{Id}$. Then $\mathop{\blacklozenge}\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p} = \sum_{k=1}^{p} \alpha_k g_k \circ L_k$.

Example 6 Suppose that Assumption 1 is in force and that, for every $k \in \{1,\ldots,p\}$, $\mathcal{G}_k = \mathcal{H}$ and $L_k = \operatorname{Id}$. Then it follows from results of [9] that Definition 3 reduces to the *proximal average* of $(g_k)_{1\leqslant k\leqslant p}$, namely,

$$\mathop{\blacklozenge}\mathsf{M}(\operatorname{Id}, g_k)_{1\leqslant k\leqslant p} = \left(\sum_{k=1}^{p} \alpha_k (g_k^* \,\square\, \mathscr{Q}_{\mathcal{G}_k})\right)^* - \mathscr{Q}_{\mathcal{H}}. \tag{8}$$

This construct has been studied in [1], [2] and applied to data analysis problems in [14], [17], [19], [21].

Example 7 Suppose that Assumption 1 is in force, that $f = 0$, and that, for every $k \in \{1,\ldots,p\}$, $g_k = \iota_{D_k}$, where D_k is a nonempty closed convex subset of $\mathcal{G}_k$. In this case (1) reduces to (2), while Definition 3 yields

$$\mathop{\blacklozenge}\mathsf{M}(L_k, \iota_{D_k})_{1\leqslant k\leqslant p} = \left(\left(\frac{1}{2}\sum_{k=1}^{p} \alpha_k d_{D_k}^2 \circ L_k\right)^* - \mathscr{Q}_{\mathcal{H}}\right)^*. \tag{9}$$

By Proposition 4(iii), the set of minimizers of (9) coincides with that of the function $x \mapsto \sum_{k=1}^{p} \alpha_k d_{D_k}^2(L_k x)$, which has been used in least-squares relaxation of inconsistent feasibility problems [4], [8]. This robust behavior can be established for more general settings beyond convex feasibility [9].

Example 8 Let $\{\mathcal{V}, \mathcal{E}\}$ be an undirected graph, where $\mathcal{V} = \{1,\ldots,M\}$ is the set of nodes and $\mathcal{E}$ is the set of edges. For every edge $(i,j) \in \mathcal{E}$, let $\mathcal{G}_{ij}$ be a real Hilbert space, let $\alpha_{ij} \in]0,1]$, let $L_{ij}\colon \mathcal{H} \to \mathcal{G}_{ij}$ be linear and bounded with $\|L_{ij}\| \leqslant 1$, and let $g_{ij} \in \Gamma_0(\mathcal{G}_{ij})$. Suppose that $\sum_{(i,j)\in\mathcal{E}} \alpha_{ij} = 1$. In the spirit of existing graph regularizers, one can consider the abstract loss function

$$\sum_{(i,j)\in\mathcal{E}} \alpha_{ij} g_{ij} \circ L_{ij}, \tag{10}$$

which is based on a standard composite average. The corresponding proximal comixture is

$$\mathop{\blacklozenge}\mathsf{M}(L_{ij}, g_{ij})_{(i,j)\in\mathcal{E}} \tag{11}$$

which, by Proposition 4(ii), has an explicit proximity operator. The setting of [15] in the context of feature selection utilizes (10) with $\mathcal{H} = \mathbb{R}^N$ and, for every $(i,j) \in \mathcal{E}$, $\mathcal{G}_{ij} = \mathbb{R}^N$, $L_{ij} = \operatorname{Id}$, and $g_{ij}\colon (\xi_l)_{1\leqslant l\leqslant N} \mapsto |\xi_i - \xi_j|$. By contrast, [21] applied implicitly the comixture (11) in the form of the proximal average (8) in this specific scenario. More generally, (11) can be considered as an alternative to (10) as a graph-based regularizer.

IV. The proximal comixture minimization problem

We consider the following alternative to Problem 2.

Problem 9 Under Assumption 1, the task is to

$$\operatorname*{minimize}_{x\in\mathcal{H}}\ f(x) + \Big(\mathop{\blacklozenge}\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p}\Big)(x), \tag{12}$$

assuming that a solution exists.

Since, by virtue of Proposition 4(ii), the proximity operator of the second term is explicit, two direct algorithms can be devised for solving (12): the Douglas–Rachford algorithm [1, Section 28.3] in general, and the forward-backward algorithm [1, Section 28.5] if f is smooth.

Proposition 10 (Douglas–Rachford) *In Problem 9, suppose that* $0 \in \operatorname{range}(\partial f + \partial(\blacklozenge\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p}))$. *Let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$ *such that* $\sum_{k\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, *let* $y_0 \in \mathcal{H}$, *and iterate*

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = y_n + \sum_{k=1}^{p} \alpha_k L_k^* \big(\operatorname{prox}_{g_k}(L_k y_n) - L_k y_n\big) \\ z_n = \operatorname{prox}_f(2x_n - y_n) \\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array} \right. \end{array} \tag{13}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 9.*

Proposition 11 (Forward-backward) *In Problem 9, suppose that* f *is differentiable on* $\mathcal{H}$ *with a* β*-Lipschitzian gradient, where* $\beta \in]0,2[$, *let* $x_0 \in \mathcal{H}$, *and iterate*

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \nabla f(x_n) \\ x_{n+1} = y_n + \sum_{k=1}^{p} \alpha_k L_k^* \big(\operatorname{prox}_{g_k}(L_k y_n) - L_k y_n\big). \end{array} \right. \end{array} \tag{14}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 9.*

Let us add that, as shown in [5], inertia can be added in Proposition 11 to obtain optimal rates of convergence for the values of the objective in (12). To solve Problem 2, we shall use the Condat–Vũ algorithm [12], [20]. Unlike (13) and (14), which split two functions and store two variables, it splits $p+1$ functions and stores $p+1$ variables at each iteration. The same holds true for other algorithms for solving Problem 2 [6], [11], [13].

Proposition 12 *In Problem 2, suppose that* $0 \in \operatorname{range}(\partial f + \sum_{k=1}^{p} \alpha_k L_k^* \circ \partial g_k \circ L_k)$. *Let* τ *and* σ *be in* $]0,+\infty[$ *and, for every* $k \in \{1,\ldots,p\}$, *let* $v_{k,0}^* \in \mathcal{G}_k$. *Suppose that* $\tau\sigma\sum_{k=1}^{p}\|L_k\|^2 < 1$, *let* $x_0 \in \mathcal{H}$, *and iterate*

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \tau\sum_{k=1}^{p} L_k^* v_{k,n}^* \\ x_{n+1} = \operatorname{prox}_{\tau f} y_n \\ z_n = 2x_{n+1} - x_n \\ \text{for every } k \in \{1,\ldots,p\} \\ \left\lfloor \begin{array}{l} w_{k,n} = v_{k,n}^* + \sigma L_k z_n \\ v_{k,n+1}^* = w_{k,n} - \sigma \operatorname{prox}_{\alpha_k g_k/\sigma}(w_{k,n}/\sigma). \end{array} \right. \end{array} \right. \end{array} \tag{15}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem 2.*

V. Applications

Since the algorithms have essentially the same computational load per iteration, we compare them in terms of error versus iteration number. The results of the three experiments conducted below consistently support the fact that the proximal comixture models lead to reliable solutions and faster algorithms which, in addition, are much less demanding in terms of memory requirements.

A. Experiment 1: Multiview image reconstruction from partial diffraction data

We consider the problem of reconstructing the image $\overline{x} \in C = [0,255]^N$ ($N = 256^2$) of Fig. 1(a) from a partial observation of its diffraction over some frequency range R, possibly with measurement errors [18]. To exploit this information we use the soft constraint penalty d_E associated with the set

$$E = \big\{x \in \mathbb{R}^N \mid (\forall k \in R)\ \widehat{x}(k) = \widehat{\overline{x}}(k)\big\}, \tag{16}$$

where $\widehat{x}$ denotes the two-dimensional discrete Fourier transform of x. 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. In addition, two blurred noisy observations of $\overline{x}$ are available, namely (see Fig. 1(b)–(c)) $z_1 = L_1\overline{x} + w_1$ and $z_2 = L_2\overline{x} + w_2$. Here, L_1 and L_2 model convolutional blurs with constant kernels of size 3×11 and of 7×5, respectively, and w_1 and w_2 are Gaussian white noise realizations. The blurred image-to-noise ratios are 30.1 dB and 34.6 dB.

Problem 13 In Problem 2, set $f = \iota_C$, $p = 4$, $\alpha_1 = \alpha_2 = 3/8$, $\alpha_3 = \alpha_4 = 1/8$, $g_1 = \mathfrak{h}_\rho \circ \|\cdot - z_1\|$, $g_2 = \mathfrak{h}_\rho \circ \|\cdot - z_2\|$, $g_3 = d_E$, $L_3 = \mathrm{Id}$, $g_4 = \sqrt{8}\|\cdot\|_1$, and $L_4 = D/\sqrt{8}$, where $\rho = 300$, $\mathfrak{h}_\rho$ is defined in (7), and $D\colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ models finite differences. The task is to

$$\begin{aligned}\underset{x\in C}{\text{minimize}}\ \ \frac{3}{8}\mathfrak{h}_\rho(\|L_1x - z_1\|) + \frac{3}{8}\mathfrak{h}_\rho(\|L_2x - z_2\|) + \\ \frac{1}{8}d_E(x) + \frac{1}{8}\big(\sqrt{8}\|L_4x\|_1\big).\end{aligned} \tag{17}$$

Problem 14 In Problem 9, define f, p, $(\alpha_k)_{1\leqslant k\leqslant p}$, $(g_k)_{1\leqslant k\leqslant p}$, and $(L_k)_{1\leqslant k\leqslant p}$ as in Problem 13, and replace (17) by

$$\underset{x\in C}{\text{minimize}}\ \ \Big(\blacklozenge\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p}\Big)(x). \tag{18}$$

We apply Propositions 12 and 10 to Problems 13 and 14, respectively, with all initial vectors set to 0. The parameters used in Propositions 12 are $\sigma = 1/(1.1\beta)$ and $\tau = 1/\beta$, where $\beta = \sqrt{\sum_{k=1}^{4}\|L_k\|^2}$, as these values gave faster converge of the algorithm. The restored images are shown in Fig. 2, while Fig. 3 illustrates the faster convergence of the proximal comixture model compared to the standard composite average.

(a) (b)

(c)

Fig. 1: (a) Original image $\overline{x}$. (b) Degraded image z_1. (c) Degraded image z_2.

(a) (b)

Fig. 2: (a) Image restored by Problem 13/Proposition 12. (b) Image restored by Problem 14/Proposition 10.

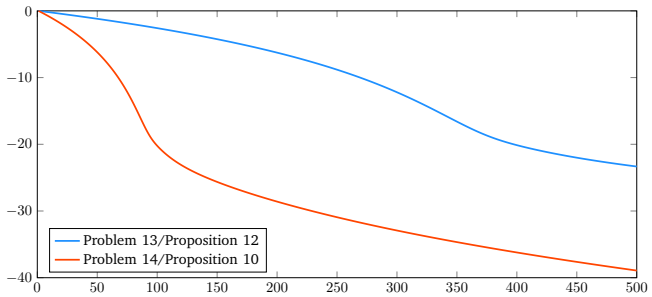


Fig. 3: Normalized error $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus iteration count in Experiment V-A.

B. Experiment 2: Image reconstruction from phase

This numerical example addresses a phase recovery problem considered in [10]. The goal is to recover the image $\overline{x} \in C = [0, 255]^N$ ($N = 512^2$) from the observation of its Fourier phase $\theta = \angle\widehat{\overline{x}}$ [16]. The original image $\overline{x}$ is shown in Fig. 4(a). The problem is modeled as a convex feasibility problem with the following constraint sets.

- Phase: $C_1 = \big\{x \in \mathbb{R}^N \mid \angle\,\widehat{x} = \theta\big\}$.
- Mean pixel value: $C_2 = \big\{x \in \mathbb{R}^N \mid \langle x \,|\, 1\rangle = \eta\big\}$.
- Upper bound on the norm of the gradient: $Dx/\sqrt{8} \in C_3$, where $C_3 = \big\{y \in \mathbb{R}^N \times \mathbb{R}^N \mid \|y\|_2 \leqslant \rho\big\}$ and D is defined as in Problem 13.
- Proximity to the reference image z of Fig. 4(b): $C_4 = \big\{x \in \mathbb{R}^N \mid \|x - z\|_2 \leqslant \xi\big\}$. The image z 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) and the addition of a local high intensity noise on a rectangular area around the right eye.

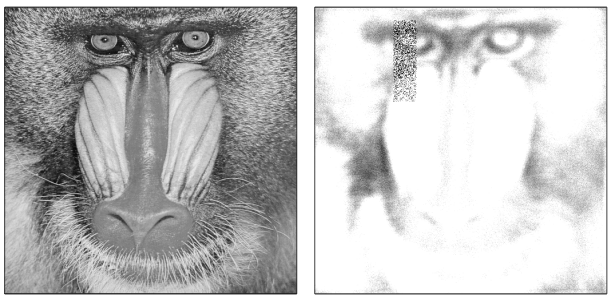
Because of inaccuracies in the values θ, η, ρ, and ξ, this problem is inconsistent and it is relaxed as follows.

Problem 15 In Problem 2, set $f = \iota_C$, $p = 4$, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1/4$, $g_1 = \mathfrak{h}_{\rho_1} \circ d_{C_1}$, $L_1 = \mathrm{Id}$, $g_2 = \mathfrak{h}_{\rho_2} \circ d_{C_2}$, $L_2 = \mathrm{Id}$, $g_3 = \mathfrak{h}_{\rho_3} \circ d_{C_3}$, $L_3 = D/\sqrt{8}$, $g_4 = \mathfrak{h}_{\rho_4} \circ d_{C_4}$, and $L_4 = \mathrm{Id}$, where $\rho_1 = \rho_2 = \rho_3 = 3000$, $\rho_4 = 5000$, and $\mathfrak{h}_\rho$ is defined in (7). The task is to

$$\underset{x \in C}{\text{minimize}} \;\; \frac{1}{4}\mathfrak{h}_{\rho_1}(d_{C_1}(x)) + \frac{1}{4}\mathfrak{h}_{\rho_2}(d_{C_2}(x)) + \frac{1}{4}\mathfrak{h}_{\rho_3}(d_{C_3}(L_3 x)) + \frac{1}{4}\mathfrak{h}_{\rho_4}(d_{C_4}(x)). \quad (19)$$

Problem 16 In Problem 9, define f, p, $(\alpha_k)_{1\leqslant k\leqslant p}$, $(g_k)_{1\leqslant k\leqslant p}$ and $(L_k)_{1\leqslant k\leqslant p}$ as in Problem 15, and replace (19) by

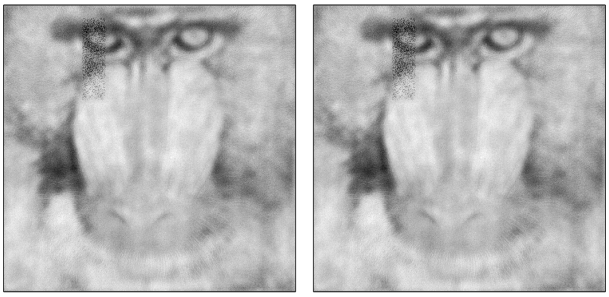
$$\underset{x \in C}{\text{minimize}} \;\; \Big(\blacklozenge\mathsf{M}(L_k, g_k)_{1\leqslant k\leqslant p}\Big)(x). \quad (20)$$



(a) (b)

Fig. 4: (a) Original image $\overline{x}$. (b) Reference image z.

We apply Propositions 12 and 10 to Problems 15 and 16, respectively, with all initial vectors set to 0. The parameters used in Propositions 12 are $\sigma = 1/(1.1\beta)$ and $\tau = 1/\beta$, where $\beta = \sqrt{\sum_{k=1}^{4} \|L_k\|^2}$, as these values gave faster convergence of the algorithm. The restored images are shown in Fig. 5, while Fig. 6 illustrates the faster convergence of the proximal comixture model compared to the standard composite average.



(a) (b)

Fig. 5: (a) Image restored by Problem 15/Proposition 12. (b) Image restored by Problem 16/Proposition 10.

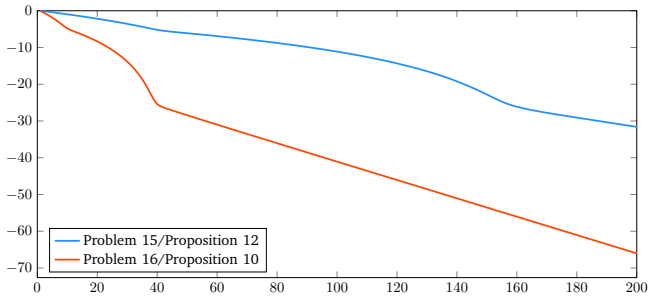


Fig. 6: Normalized error $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus iteration count in Experiment V-B.

C. Experiment 3: Overlapping group lasso

We consider the following instance of Problem 2.

Problem 17 The task is to solve the overlapping group lasso problem [7]

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \frac{1}{2}\|Ax - z\|^2 + \sum_{k=1}^{p} \alpha_k \|L_k x\|, \qquad (21)$$

where $p = 50$, $N = 2255$, $M = 2000$, $A \in \mathbb{R}^{M\times N}$ is normalized so that $\|A\| \leqslant 1$, $\overline{x} = (\overline{\xi}_j)_{1\leqslant j\leqslant N}$, where $\overline{\xi}_j = (-1)^j \exp(-(j-1)/50)$, $z = A\overline{x} + w$, where w is a realization of a Gaussian noise with zero mean and unit variance, and, for every $k \in \{1,\ldots,p\}$, $L_k\colon x \mapsto (\xi_{45(k-1)+1},\ldots,\xi_{45(k-1)+50})$ and $\alpha_k = 1/p$.

Problem 18 In Problem 17, replace (21) by

$$\underset{x\in\mathbb{R}^M}{\text{minimize}} \;\; \frac{1}{2}\|Ax - b\|^2 + \Big(\mathop{\blacklozenge}_{1\leqslant k\leqslant p} \mathsf{M}(L_k, \|\cdot\|)\Big)(x). \qquad (22)$$

While the methods converge to similar solutions, Fig. 7 shows that the proximal comixture approach yields faster convergence than that of the standard composite average.

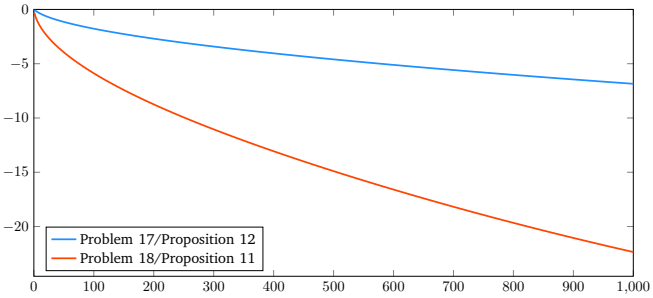


Fig. 7: Normalized error $20\log_{10}(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus iteration count in Experiment V-C.

References

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

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

[3] M. N. Bùi and P. L. Combettes, Integral resolvent and proximal mixtures, *J. Optim. Theory Appl.*, to appear.

[4] 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.

[5] A. Chambolle and C. Dossal, On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm," *J. Optim. Theory Appl.*, vol. 166, pp. 968–982, 2015.

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

[7] 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.

[8] P. L. Combettes, The convex feasibility problem in image recovery, in *Advances in Imaging and Electron Physics*, vol. 95, pp. 155–270. Academic, New York, 1996.

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

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] S. Kim and E. P. Xing, Statistical estimation of correlated genome associations to a quantitative trait network, *PLoS Genetics*, vol. 5, pp. 1–18, 2009.

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

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

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

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

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

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