A variational approach for joint image recovery-segmentation based on spatially varying generalised Gaussian models

Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet, Gabriele Scrivanti

A VARIATIONAL APPROACH FOR JOINT IMAGE RECOVERY-SEGMENTATION BASED ON SPATIALLY VARYING GENERALISED GAUSSIAN MODELS

É. CHOUZENOUX*, M.-C. CORBINEAU*†, J.-C. PESQUET*, AND G. SCRIVANTI*‡

Abstract. The joint problem of reconstruction / feature extraction is a challenging task in image processing. It consists in performing, in a joint manner, the restoration of an image and the extraction of its features. In this work, we firstly propose a novel nonsmooth and nonconvex variational formulation of the problem. For this purpose, we introduce a versatile generalised Gaussian prior whose parameters, including its exponent, are space-variant. Secondly, we design an alternating proximal-based optimisation algorithm that efficiently exploits the structure of the proposed nonconvex objective function. We also analyze the convergence of this algorithm. As shown in numerical experiments conducted on joint segmentation/deblurring tasks, the proposed method provides high-quality results.

Key words. Image recovery ; Space-variant regularisation ; Alternating minimization ; Proximal algorithm ; Block coordinate descent ; Kurdyka–Łojasiewicz property ; Variable metric ; Image segmentation ; Texture decomposition ; Ultrasound imaging

AMS subject classifications.

1. Introduction.

Variational regularisation of ill-posed inverse problems in imaging relies on the idea of searching for a solution in a well-suited space. A central role in this context is played by ℓ_p spaces with $p \in (0, \infty)$, and the power p of the corresponding norms when $p \geq 1$ [32, 39, 50, 58, 63] or seminorms when $p \in (0, 1)$ [22, 38, 73]. For every vector $u = (u_i)_{1 \leq i \leq n} \in \mathbb{R}^n$ and $p \in (0, +\infty)$, the ℓ_p (semi-)norm is denoted by $\|u\|_p = \left(\sum_{i=1}^n |u_i|^p\right)^{1/p}$. We usually omit p when $p = 2$, so that $\|\cdot\| = \|\cdot\|_2$. The case $p \in (0, 1)$ has gained rising credit, especially in the field of sparse regularisation. An extensive literature has been focused on challenging numerical tasks raised by the nonconvexity of the seminorms and the possibility to combine them with linear operators to extract salient features of the sought images [37, 42]. In [51] the more general notion of F-norm is introduced in order to establish functional analysis results on products of ℓ_{p_i}-spaces with $p_i \in (0, 2]$. For some $x = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, this amounts to studying the properties of penalties of the form $\sum_{i=1}^n |x_i|^{p_i}$, for some positive exponents $(p_i)_{1 \leq i \leq n}$. This approach offers a more flexible framework by considering a wider range of exponents than the standard ℓ_p-based regularisation. However, it extends the problem of choosing a suitable exponent p to a whole sequence of exponents $(p_i)_{1 \leq i \leq n}$. In image restoration, a related approach consists in adopting space variant regularisation models built around a Total Variation-like functional with a variable exponent $\sum_{i=1}^n \|(\nabla x)_i\|^{p_i}$ where ∇ is a discrete 2D gradient operator. The rationale is to select the set of parameters $(p_i)_{1 \leq i \leq n}$ in order to promote either edge enhancement ($p_i = 1$) or smoothing ($p_i > 1$) depending on the spatial location encoded by index i. This model was introduced in [8] and then put into practice firstly for $p_i \in [1, 2]$ in [23] and then for $p_i \in (0, 2]$ in [46]. In all of these works, the so-called space variant p-map (i.e., $(p_i)_{1 \leq i \leq n}$) is estimated offline in a preliminary step and then kept fixed throughout the optimisation procedure.

*Université Paris-Saclay, Inria, CentraleSupélec, CVN, France. (firstname.name@centralesupelec.fr)

†Research department, Preligens

‡Corresponding author

In this paper, we address the problem of joint image recovery and features extraction. Image recovery amounts to retrieving an estimate of an original image from a degraded version of it. The degradation usually corresponds to the application of a linear operator (e.g., blur, projection matrix) to the image, and the addition of a noise. Feature extraction problems arise when one wants to assign to an image a small set of parameters which can describe or identify the image itself. Image segmentation can be viewed as an example of features extraction, which consists of defining a label field on the image domain so that pixels are partitioned into a predefined number of homogeneous regions according to some specific characteristics. Texture retrieval is a second example. This task relies on the idea of assigning a set of parameters to each coefficients of the image, in some transformed space, so that the combination of all parameters defines a "signature" that represents the content of various spatial regions. Joint image recovery and feature extraction consists in performing, in a joint manner, the image recovery and the extraction of features in the sought image. A powerful and versatile approach for features extraction, that we will adopt here, assumes that the image data follow a mixture of generalised Gaussian probability distribution ($\mathcal{GGD}$) [30, 33, 74]. In our case, this leads to the minimization of a non-smooth and non-convex cost function. The $\mathcal{GGD}$ model results in a sum of weighted ℓ_{p_i}-based terms in the criterion, with general form $\sum_{i=1}^{n} \vartheta_i |x_i|^{p_i}$ with $\{\vartheta_i\}_{1\leq i\leq n} \subset [0, +\infty)$. We thus aim at jointly estimating an optimal configuration for $(\vartheta_i, p_i)_{1\leq i\leq n}$, and retrieving the image. Under an assumption of consistency within the exponents values of a given region of the features space, we indeed obtain the desired feature extraction starting from the estimated p-map.

The specific structure of the objective function we will propose suggests the use of an alternating minimisation procedure. In such an approach, one sequentially updates a subset of parameters through the resolution of an inner minimization problem, while the other parameters are assumed to be fixed. This approach has a standard form in the Block Coordinate Descent method (BCD) (also known as Gauss-Seidel algorithm) [41]. In the context of nonsmooth and nonconvex problems, the simple BCD may show instabilities [65], which resulted in an extensive construction of alternative methods that efficiently exploit the characteristics of the functions, and introduce powerful tools to improve the convergence guarantees of BCD, or overcome difficulties arising in some formulations. In this respect, a central role is played by proximal methods [26, 27]: a proximally regularised BCD (PAM) for nonconvex problems was studied in [5]; a proximal linearised method (PALM) and its inertial and stochastic versions were then proposed in [11] resp. [57] and [40]; in [35] the authors investigated the advantage of a hybrid semi-linearised scheme (SL-PAM) for the joint task of image restoration and edge detection based on a discrete version of the Mumford–Shah model. A structure-adapted version of PALM (ASAP) was designed in [53] to exploit the block-convexity of the coupling terms and the regularity of the block-separable terms arising in some practical applications such as nonnegative matrix factorisation and blind source separation. The extension to proximal mappings defined w.r.t. a variable metric was firstly introduced in [18], leading to the so-called Block Coordinate Variable Metric Forward Backward. An Inexact version and a linesearch based version of it were presented in [25] and [13], respectively. In [59] the authors introduced a Majorisation-Minimisation strategy in a Variable Metric Forward-Backward algorithm to tackle the challenging task of computing the proximity operator of composite functions. We refer to [14] for an in-depth analysis of how to introduce a variable metric into first-order methods. To conclude this brief overview, let us also mention generalisations to proximal mappings defined according

to Bregman distances proposed in [2] and [47] , which extend to the block coordinate case the attempt in [9, 12] to relax the common assumption of Lipschitz continuity on the gradient of the smooth term.

In the problem formulation we will be interested in, the objective function includes a quadratic term, hence Lipschitz differentiable, that is restricted to a single block of parameters. This feature makes the related subproblem well-suited for a splitting procedure that involves an explicit gradient step with respect to this term. None of the aforementioned methods provide a proper framework for this purpose. In order to exploit the described particular structure, we thus propose a new BCD method that mixes standard and linearised proximal regularisations on the different blocks (as in SL-PAM). The novelty lies in the fact that we include a preconditioned and structure-adapted linearised step to obtain more efficient and faster proximal computations (as in ASAP). We refer to the proposed method as the Preconditioned Semi-Linearised Structure Adapted Proximal Alternating Minimisation (P-SASL-PAM). We investigate the convergence properties for this algorithm based on the framework designed in [6]. Under analytical assumptions on the objective function, we show the global convergence toward a critical point of any sequence generated by the proposed method. Then, we explicit the use of this method in our problem of image recovery and feature extraction. The performance of the approach is illustrated by means of examples in the field of ultrasound imaging, in which we also show quantitative comparisons with state-of-the art-methods for the joint deconvolution-segmentation task.

The contributions of this work are (i) the proposition of an original variational model for the joint image recovery and features extraction problem; (ii) the design of a new block coordinate descent algorithm to address the resulting minimisation problem; (iii) the convergence analysis of this scheme based on [6]; (iv) the illustration of the performance of the proposed method through two numerical examples in the field of image processing.

The paper is organised as follows. In Section 2 we introduce the degradation model and report our derivation of the objective function for image recovery and feature extraction, starting from statistical assumptions on the data. In Section 3, we describe the proposed P-SASL-PAM method to address a general non-smooth non-convex optimization problem; secondly we show that the proposed method converges globally, in the sense that the whole generated sequence converges to a (local) minimum. The application of the P-SASL-PAM method to the joint reconstruction-segmentation problem is described in Section 4. Some illustrative numerical results are shown in Section 5. Conclusions are drawn in Section 6

2. Model Formulation. In this section, we describe the construction of the objective function associated to the joint reconstruction-feature extraction problem. After defining the degradation model, we report the Bayesian model that is reminiscent from the one considered in [30, 74] in the context of ultrasound imaging. Then, we describe the procedure that leads us to the definition of our addressed optimization problem.

2.1. Observation Model. Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be respectively the vectorized sought-for solution and the observed data, which are assumed to be related according to the following model

$$y = Kx + \omega, \tag{2.1}$$

where $K \in \mathbb{R}^{m\times n}$ is a linear operator, and $\omega \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_m)$, i.e. the normal distribution with zero mean and covariance matrix $\sigma^2\mathbb{I}_m$ with $\sigma > 0$ and $\mathbb{I}_m$ states for the $m \times m$ identity matrix. We further assume that x can be characterised by a finite set of k *features* that are defined in a suitable space, where the data are described by a simple model relying on a small number of parameters. The Generalised Gaussian Distribution ($\mathcal{GGD}$)

$$(\forall t \in \mathbb{R}) \quad \mathsf{p}(t; p, \alpha) = \frac{1}{2\alpha^{1/p}\Gamma\left(1+\frac{1}{p}\right)} \exp\left(-\frac{|t|^p}{\alpha}\right), \quad (p, \alpha) \in (0, +\infty)^2 \tag{2.2}$$

has shown to be a suitable and flexible tool for this purpose [30, 33, 74]. Each feature can be identified by a pair (p_j, α_j) for $j \in \{1, \dots, k\}$, where parameter p is proportional to the decay rate of the tail of the probability density function (PDF) and parameter α models the width of the peak of the PFD. Taking into account the role that p and α play in the definition of the PDF profile, these two parameters are generally referred to as *shape* and *scale* parameter.

Assuming that K and σ are known, the task we address in this work is to jointly retrieve x (reconstruction) and obtain a good representation of its features through an estimation of the underlying model parameters (p_j, α_j) for $j \in \{1, \dots, k\}$ (feature extraction). Starting from a similar statistical model as the one considered in [30, 74], we infer a continuous variational framework which does not rely on the *a priori* knowledge of the exact number of features k. We derive this model by performing a *Maximum a Posteriori* estimation, which allows us to formulate the Joint Image Reconstruction and Feature Extraction task as a nonsmooth nonconvex optimisation problem involving a coupling term and a block-coordinate separable one.

2.2. Bayesian Model. From (2.1), we derive the following likelihood

$$\mathsf{p}(y|x, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|y - Kx\|^2}{2\sigma^2}\right). \tag{2.3}$$

Assuming then that the components of x are independent conditionally to the knowledge of their feature class, we define x as a mixture of $\mathcal{GGD}$s

$$\mathsf{p}(x|p, \alpha) = \prod_{j=1}^{k} \frac{1}{\left(2\alpha_j^{1/p_j}\Gamma\left(1+\frac{1}{p_j}\right)\right)^{N_j}} \exp\left(-\frac{\|\overline{x}_j\|_{p_j}^{p_j}}{\alpha_j}\right). \tag{2.4}$$

Hereabove, for every $u \in \mathbb{R}^n$ and a feature labels set $j \in \{1, \dots, k\}$, we define $\overline{u}_j \in \mathbb{R}^{N_j}$ as the vector containing only the N_j components of u that belong to the j-th feature. The shape parameters are assumed to be uniformly distributed on a certain interval $[a, b] \subset \mathbb{R}^{+*}$, while uninformative Jeffreys priors are assigned to the scale parameters:

$$\mathsf{p}(p) = \prod_{j=1}^{k} \frac{1}{b-a} \mathsf{I}_{[a,b]}(p_j), \tag{2.5}$$

$$\mathsf{p}(\alpha) = \prod_{j=1}^{k} \frac{1}{\alpha_j} \mathsf{I}_{\mathbb{R}_+}(\alpha_j). \tag{2.6}$$

Hereabove, I_S represents the characteristic function of some subset $S \subset \mathbb{R}$, which is equal to 1 over S, and 0 elsewhere.

2.3. Variational Model. In order to avoid to define *a priori* the number of features, we regularise the problem by considering the 2D Total Variation (TV) of the $\mathcal{GGD}$ parameters $(p, \alpha) \in (0, +\infty)^n \times (0, +\infty)^n$. The idea of using Total Variation to define a segmentation procedure is studied in [15, 16, 17, 19, 20, 55] by virtue of the co-area formula: the authors propose to replace the boundary information term of the Mumford-Shah (MS) functional [52] with the TV convex integral term. This choice yields a nontight convexification of the MS model that does not require to set the number of segments in advance. The overall segmentation procedure is then built upon two steps: the first one consists of obtaining a smooth version of the given image that is adapted to segmentation by minimising the proposed functional with convex methods; the second step consists of partitioning the obtained solution into the desired number of segments, by *e.g.* defining the thresholds with Otsu's method [54] or the k-means algorithm. The strength of our approach is that the second step (*i.e.* the actual segmentation step) is independent from the first one, hence it is possible to set the number of segments (i.e., labels) without solving the optimisation problem again.

In the considered model, the introduction of a TV prior leads to a minimization problem that is nonconvex w.r.t α. Preliminary experimental results suggested the use of the following reparameterisation for the scale parameter. Let $\beta = (\beta_i)_{1\leq i\leq n} \in \mathbb{R}^n$ be such that for every $i \in \{1, \dots, n\}$,

$$\beta_i = \frac{1}{p_i} \ln \alpha_i, \tag{2.7}$$

and let us choose for this new variable a normal prior with zero-mean and standard deviation σ_β. Replacing α with β and introducing the TV regularisation (weighted by the regularisation parameters $\lambda > 0$ and $\zeta > 0$) leads to the following re-parameterization of distributions (2.4)-(2.6):

$$\mathsf{p}(x|p, \beta) = \prod_{i=1}^{n} \frac{1}{2\exp(\beta_i)\Gamma\left(1 + \frac{1}{p_i}\right)} \exp\left(-|x_i|^{p_i} \exp(-p_i\beta_i)\right) \tag{2.8}$$

$$\mathsf{p}(p) = \exp(-\lambda \mathrm{TV}(p)) \prod_{i=1}^{n} \frac{1}{b-a} \mathsf{I}_{[a,b]}(p_i) \tag{2.9}$$

$$\mathsf{p}(\beta) = \exp(-\zeta \mathrm{TV}(\beta)) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\Big(-\frac{\beta_i^2}{2\sigma_\beta^2}\Big). \tag{2.10}$$

The joint posterior distribution is determined as follows:

$$\begin{aligned} \mathsf{p}(x, p, \beta|y) &\propto \mathsf{p}(y|x, p, \beta)\mathsf{p}(x, p, \beta) \\ &\propto \mathsf{p}(y|x, p, \beta)\mathsf{p}(x|p, \beta)\mathsf{p}(p)\mathsf{p}(\beta). \end{aligned} \tag{2.11}$$

Let us take the negative logarithm of (2.11), then computing the Maximum a Posteriori estimates (i.e., maximising the joint posterior distribution) is equivalent to the following optimization problem, which we refer to as the Joint Image Reconstruction

and Features Extraction Problem:

$$\underset{(x,p,\beta)\in\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n}{\text{minimize}} \quad \Theta(x,p,\beta) = \frac{1}{2\sigma^2}\|y-Kx\|^2 + \sum_{i=1}^{n}\left(|x_i|^{p_i}e^{-p_i\beta_i} + \ln\Gamma(1+\frac{1}{p_i}) + \iota_{[a,b]}(p_i) + \beta_i + \frac{\beta_i^2}{2\sigma_\beta^2}\right) + \lambda\mathrm{TV}(p) + \zeta\mathrm{TV}(\beta). \tag{2.12}$$

We notice that, when restricted to variable x for a given set of parameters (p,β), the above minimisation problem boils down to the flexible sparse regularisation model

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2\sigma^2}\|y-Kx\|^2 + \sum_{i=1}^{n}|x_i|^{p_i}e^{-p_i\beta_i}, \tag{2.13}$$

where the contribution of the ℓ_{p_i} regularisation term is itself weighted in a space varying fashion.

Function Θ in (2.12) is nonsmooth and nonconvex. It reads as the sum of a coupling term and three block-separable terms. In particular, the block-separable data-fit term relative to x is quadratic, and hence has a Lipschitz continuous gradient. Our proposed algorithm aims at leveraging this property, which is generally not accounted for by other BCD methods. To this aim, we exploit a hybrid scheme that involves both standard and linearised proximal steps. The details about the proposed method are presented in the next section.

3. Preconditioned Structure Adapted Semi-Linearised Proximal Alternating Minimisation (P-SASL-PAM).

In this section, we introduce a BCD-based method to address a class of sophisticated optimization problems including (2.12) as a special case. We start the section by useful preliminaries about subdifferential calculus. Then, we present the *Kurdyka-Łojasiewicz property*, which plays a prominent role in the convergence analysis of BCD methods in a nonconvex setting. Finally, we define problem (3.4), itself generalising (2.12), for which we derive our proposed BCD-based algorithm, and show its convergence properties. The so-called Preconditioned Structure Adapted Semi-Linearised Proximal Alternating Minimisation (P-SASL-PAM) approach mixes both standard and preconditioned linearised proximal regularisation on the different coordinate blocks of the criterion.

3.1. Subdifferential Calculus.

Let us now recall some definitions and elements of subdifferential calculus that will be useful in the upcoming sections. For a proper and lower semicontinuous function $f:\mathbb{R}^n\to(-\infty,\infty]$, the domain of f is defined as

$$\operatorname{dom} f \ = \{u\in\mathbb{R}^n \mid f(u) < +\infty\}.$$

Firstly, we recall the notion of subgradients and subdifferential for convex functions.

DEFINITION 3.1 (Subgradient of a convex function). *Let $f:\mathbb{R}^n\to(-\infty,\infty]$ be a proper convex lower semicontinuous function. The subdifferential $\partial f(u^+)$ of f at $u^+\in\mathbb{R}^n$ is the set of all vectors $v\in\mathbb{R}^n$, known as subgradients of f at u^+, such that*

$$\forall u\in\mathbb{R}^n \quad f(u)\geq f(u^+)+\langle v, u-u^+\rangle.$$

If $u^+\notin\operatorname{dom} f$, then $\partial f(u^+)=\varnothing$.

Secondly, we consider the more general notion of (limiting)-subdifferential for non necessarily convex functions, as proposed in [61, Definition 8.3].

DEFINITION 3.2 (Limiting Subdifferential). *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper and lower semicontinuous function. For a vector $u^+ \in \mathbb{R}^n$,*

- *the Fréchet subdifferential of f at u^+, written as $\hat{\partial} f(u^+)$, is the set of all vectors $v \in \mathbb{R}^n$ such that*

$$(\forall u \in \mathbb{R}^n) \;\; f(u) \geq f(u^+) + \langle v, u - u^+ \rangle + o(\|u - u^+\|);$$

  if $u^+ \notin \mathrm{dom}\, f$, then $\hat{\partial} f(u^+) = \varnothing$;
- *the limiting-subdifferential of f at u^+, denoted by $\partial f(u^+)$, is defined as*

$$\partial f(u^+) = \left\{ u \in \mathbb{R}^n \mid \exists\, u^k \to u^+, f(u^k) \to f(u^+), v^k \to v, v^k \in \hat{\partial} f(u^k) \right\}.$$

If f is lower semicontinuous and convex, then the three previous notions of subdifferentiality are equivalent, i.e. $\hat{\partial} f(u^+) = \partial f(u^+)$. If f is differentiable, then $\partial f(u^+) = \{\nabla f(u^+)\}$. Now it is possible to formalise the notion of critical point for a general function:

DEFINITION 3.3 (Critical point). *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper function. A point $u^* \in \mathbb{R}^n$ is said to be a critical (or stationary) point for f if $0 \in \partial f(u^*)$.*

Eventually, we define the notion of proximal map relative to the norm induced by a positive definite matrix.

DEFINITION 3.4. *Let $\mathcal{S}_n$ be the set of symmetric and positive definite matrices in $\mathbb{R}^{n \times n}$. For a matrix $A \in \mathcal{S}_n$, the weighted ℓ_2-norm induced by A is defined as*

$$(\forall u \in \mathbb{R}^n) \quad \|u\|_A = (u^\top A u)^{1/2}. \tag{3.1}$$

DEFINITION 3.5. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper and lower semicontinuous function, let $A \in \mathcal{S}_n$ and $u^+ \in \mathbb{R}^n$. The proximity operator of function f at u^+ with respect to the norm induced by A is defined as*

$$\mathrm{prox}_f^A(u^+) = \mathrm{argmin}_{u \in \mathbb{R}^n} \left(\frac{1}{2} \|u - u^+\|_A^2 + f(u) \right). \tag{3.2}$$

Note that $\mathrm{prox}_f^A(u^+)$, as defined above, can be the empty set. It is nonempty for every $u^+ \in \mathbb{R}^n$, if f is lower-bounded by an affine function. In addition, it reduces to a single-valued operator when f is convex.

3.2. The KŁ-Property. Most of the works related to BCD-based algorithms rely on the framework developed by Attouch, Bolte, and Svaiter in their seminal paper [6] in order to prove the convergence of block alternating strategies for nonsmooth and nonconvex problems. A fundamental assumption in [6] is that the objective function satisfies the *Kurdyka-Łojasiewicz* (KŁ) property [45, 48, 49]. We recall the definition of this property as it was given in [11]. Let $\eta \in (0, +\infty]$ and denote by Φ_h the class of concave continuous functions $\varphi : [0, +\infty) \to \mathbb{R}_+$ satisfying the following conditions:

(i) $\varphi(0) = 0$;
(ii) φ is $\mathcal{C}^1$ on $(0, \eta)$ and continuous at 0;
(iii) for every $s \in (0, \eta)$, $\varphi'(s) > 0$.

For any subset $S \subset \mathbb{R}^n$ and any point $u^+ \in \mathbb{R}^n$, the distance from u^+ to S is defined by

$$\operatorname{dist}(u^+, S) = \inf_{u \in S} \|u^+ - u\|$$

with $\operatorname{dist}(u^+, \varnothing) = +\infty$.

DEFINITION 3.6 (KŁ property). *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper and lower semicontinuous function.*

(i) Function f is said to satisfy the Kurdyka-Łojasiewicz property at $u^+ \in \operatorname{dom} \partial f$ if there exist $\eta \in (0, +\infty]$, a neighbourhood U of u^+, and a function $\varphi \in \Phi_\eta$ such that, for every $u \in U$,

$$f(u^+) < f(u) < f(u^+) + \eta \quad \Rightarrow \quad \varphi'(f(u) - f(u^+)) \operatorname{dist}(0, \partial f(u)) \geq 1. \tag{3.3}$$

(ii) Function f is said to be a KŁ function if it satisfies the KŁ property at each point of $\operatorname{dom} \partial f$.

3.3. Proposed Algorithm. Consider the general problem

$$\underset{(x,p,\beta) \in (\mathbb{R}^n)^3}{\text{minimize}} \ \Big(\theta(x, p, \beta) = q(x, p, \beta) + f(x) + g(p) + h(\beta) \Big) \tag{3.4}$$

associated to the following set of assumptions:

ASSUMPTION 1.

(i) Function $q : (\mathbb{R}^n)^3 \to \mathbb{R}$ is bounded from below and differentiable with Lipschitz continuous gradient on bounded subsets of $(\mathbb{R}^n)^3$. In other words, for every bounded subsets S of $(\mathbb{R}^n)^3$, there exists $L_{q,S} > 0$ such that, for every $(x, p, \beta) \in S$ and $(x^+, p^+, \beta^+) \in S$,

$$\begin{aligned} &\|\nabla q(x, p, \beta) - \nabla q(x^+, p^+, \beta^+)\| \\ &\qquad \leq L_{q,S}(\|x - x^+\|^2 + \|p - p^+\|^2 + \|\beta - \beta^+\|^2)^{1/2}. \end{aligned} \tag{3.5}$$

(ii) Function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable with globally Lipschitz continuous gradient of constant $L_f > 0$, and is bounded from below.

(iii) Functions $g : \mathbb{R}^n \to (-\infty, +\infty]$ and $h : \mathbb{R}^n \to (-\infty, +\infty]$ are proper, lower semicontinuous and bounded from below.

(iv) θ is a KŁ function.

We propose a block alternating algorithm to solve problem (3.4) which sequentially updates one of the three coordinate blocks (x, p, β) involved in function θ, through proximal and gradient steps. This yields Algorithm 3.1, that we call P-SASL-PAM.

Algorithm 3.1 P-SASL-PAM

Initialize x^0, p^0 and β^0
Set $A \in \mathcal{S}_n$
Set $\gamma_1 \in (0, 1)$, $\gamma_2 > 0$, $\gamma_3 > 0$
For $\ell = 0, 1, \ldots$

$$x^{\ell+1} \in \operatorname{prox}^A_{\gamma_1 q(\cdot, p^\ell, \beta^\ell)}(x^\ell - \gamma_1 A^{-1} \nabla f(x^\ell)) \tag{3.6}$$

$$p^{\ell+1} \in \operatorname{prox}_{\gamma_2 \theta(x^{\ell+1}, \cdot, \beta^\ell)}(p^\ell) \tag{3.7}$$

$$\beta^{\ell+1} \in \operatorname{prox}_{\gamma_3 \theta(x^{\ell+1}, p^{\ell+1}, \cdot)}(\beta^\ell) \tag{3.8}$$

Our P-SASL-PAM approach is inspired from the rich literature on proximal versions of BCD. We refer in particular to SLPAM [35] and ASAP [53] since our algorithm mixes both standard and linearised proximal regularisation on the coordinate blocks as in the former work, while inverting the splitting in order to gain more efficient proximal computations as in the latter. More precisely, we took advantage of the differentiability assumption on f to perform a linearised step for the update of variable x, while p and β are updated according to a standard proximal step. In addition, in order to accelerate the convergence, we introduced a preconditioned version of the linearised step which relies on the variable metric forward-backward strategy from [24]. As in [24], the preconditioning matrix $A \in \mathcal{S}_n$ is set so as to fulfill the following *majorisation condition*:

ASSUMPTION 2.

(i) The quadratic function defined, for every $x^+ \in \mathbb{R}^n$, as

$$(\forall x \in \mathbb{R}^n) \quad \phi(x, x^+) = f(x^+) + (x - x^+)^\top \nabla f(x^+) + \frac{1}{2}\|x^+ - x\|_A^2 \tag{3.9}$$

is a majorant function of f at x^+, i.e.

$$(\forall x \in \mathbb{R}^n) \quad f(x) \leq \phi(x, x^+). \tag{3.10}$$

(ii) There exist $(\underline{\nu}, \overline{\nu}) \in (0, +\infty)^2$ such that

$$\underline{\nu}\mathbb{I}_n \preceq A \preceq \overline{\nu}\mathbb{I}_n. \tag{3.11}$$

Remark that, since f satisfies Assumption 1, the *Descent Lemma* [7, Proposition A.24] applies, yielding

$$(\forall (x, x^+) \in \mathbb{R}^n \times \mathbb{R}^n) \quad f(x) \leq f(x^+) + (x - x^+)^\top \nabla f(x^+) + \frac{L_f}{2}\|x^+ - x\|^2.$$

This guarantees that the preconditioning matrix $A = L_f \mathbb{I}_n$ satisfies Assumption 2, with $\underline{\nu} = \overline{\nu} = L_f$. Apart from this simple choice for matrix A, more sophisticated construction strategies have been studied in the literature [24, 34, 43].

3.4. Convergence analysis. In this subsection, we provide some technical results regarding the sequences $(z^\ell)_{\ell \in \mathbb{N}} = \big((x^\ell, p^\ell, \beta^\ell)\big)_{\ell \in \mathbb{N}}$ and $\big(\theta(z^\ell)\big)_{\ell \in \mathbb{N}}$ generated by Algorithm (3.1) that are instrumental to prove the convergence of the proposed method. Our proof relies on the general strategy designed in [6] which is based on three main ingredients: firstly a *sufficient decrease property*, secondly an *inexact optimality condition*, and finally the *Kurdyka-Łojasiewicz property*. On the one hand, this last property does not depend on the chosen algorithm, but only on the function at hand. In our framework, it is ensured by Assumption 1(iv). On the other hand, the first two properties, only related to the design of the algorithm itself, are expressed by Lemmas 3.7 and 3.9.

LEMMA 3.7 (Sufficient decrease and finite length). *Let $(z^\ell)_{\ell \in \mathbb{N}}$ be a sequence generated by Algorithm 3.1. Then, under Assumptions 1 and 2,*

i) there exists $\mu \in (0, +\infty)$ such that for every $\ell \in \mathbb{N}$,

$$\theta(z^{\ell+1}) \leq \theta(z^\ell) - \frac{\mu}{2}\|z^{\ell+1} - z^\ell\|^2. \tag{3.12}$$

ii) $\sum_{\ell=0}^{+\infty} \|z^{\ell+1} - z^\ell\|^2 < +\infty$ *and* $\lim_{\ell\to+\infty} z^{\ell+1} - z^\ell = 0$.

Proof. Let $\ell \in \mathbb{N}$.

i) Based on the variational definition of the proximity operator induced by the weighted norm, $x^{\ell+1}$ belongs to the set given by

$$\begin{aligned}(3.13)\quad &\operatorname{prox}^A_{\gamma_1 q(\cdot,p^\ell,\beta^\ell)}(x^\ell - \gamma_1 A^{-1}\nabla f(x^\ell))\\ &\qquad = \operatorname{argmin}_{u\in\mathbb{R}^n}\left\{q(u,p^\ell,\beta^\ell) + \frac{1}{2\gamma_1}\|u-x^\ell\|_A^2 + \langle \nabla f(x^\ell), u - x^\ell\rangle\right\}.\end{aligned}$$

Hence,

$$(3.14)\quad q(x^{\ell+1},p^\ell,\beta^\ell) + \frac{1}{2\gamma_1}\|x^{\ell+1}-x^\ell\|_A^2 + \langle \nabla f(x^\ell), x^{\ell+1}-x^\ell\rangle \le q(x^\ell,p^\ell,\beta^\ell).$$

Moreover, the majorisation property (3.10) leads to

$$(3.15)\qquad f(x^{\ell+1}) \le f(x^\ell) + \langle x^{\ell+1}-x^\ell, \nabla f(x^\ell)\rangle + \frac{1}{2}\|x^{\ell+1}-x^\ell\|_A^2.$$

Adding the quantity $f(x^\ell) + \frac{1}{2}\|x^{\ell+1}-x^\ell\|_A^2$ on both sides of (3.14) allows us to exploit (3.15) together with (3.11), to obtain

$$(3.16)\quad q(x^{\ell+1},p^\ell,\beta^\ell) + f(x^{\ell+1}) + \frac{\nu}{2}\left(\frac{1}{\gamma_1}-1\right)\|x^{\ell+1}-x^\ell\|^2 \le q(x^\ell,p^\ell,\beta^\ell) + f(x^\ell),$$

where we have used the fact that, since $\gamma_1 \in (0,1)$, $\frac{\nu}{2}\left(\frac{1}{\gamma_1}-1\right) > 0$.

Then, the variational definition of the proximal steps on functions p and β implies that

$$(3.17)\quad \frac{1}{2\gamma_2}\|p^{\ell+1}-p^\ell\|^2 + g(p^{\ell+1}) + q(x^{\ell+1},p^{\ell+1},\beta^\ell) \le g(p^\ell) + q(x^{\ell+1},p^\ell,\beta^\ell),$$

$$(3.18)\quad \frac{1}{2\gamma_3}\|\beta^{\ell+1}-\beta^\ell\|^2 + h(\beta^{\ell+1}) + q(x^{\ell+1},p^{\ell+1},\beta^{\ell+1}) \le h(\beta^\ell) + q(x^{\ell+1},p^{\ell+1},\beta^\ell).$$

In conclusion, combining (3.16), (3.17), and (3.18) yields

$$\begin{aligned}&\frac{\nu}{2}\left(\frac{1}{\gamma_1}-1\right)\|x^{\ell+1}-x^\ell\|^2 + \frac{1}{2\gamma_2}\|p^{\ell+1}-p^\ell\|^2 + \frac{1}{2\gamma_3}\|\beta^{\ell+1}-\beta^\ell\|^2\\ &\qquad\qquad \le \theta(x^\ell,p^\ell,\beta^\ell) - \theta(x^{\ell+1},p^{\ell+1},\beta^{\ell+1}).\end{aligned}\quad (3.19)$$

Thus, by setting $(z^\ell)_{\ell\in\mathbb{N}} = \big((x^\ell,p^\ell,\beta^\ell)\big)_{\ell\in\mathbb{N}}$ and defining the positive constant $\mu = \min\{\underline{\nu}(\frac{1}{\gamma_1}-1), \frac{1}{\gamma_2}, \frac{1}{\gamma_3}\}$, we get (3.12).

ii) From (3.12), it follows that the sequence $(\theta(z^\ell))_{\ell\in\mathbb{N}}$ is nonincreasing. Since function θ is assumed to be bounded from below, this sequence converges to some

real number $\underline{\theta}$. We have then, for every integer N,

$$\sum_{\ell=0}^{N} \|z^{\ell} - z^{\ell+1}\|^2 \leq \frac{1}{\mu} \sum_{\ell=0}^{N} \big(\theta(z^{\ell}) - \theta(z^{\ell+1})\big) \tag{3.20}$$

$$= \frac{1}{\mu}(\theta(z^0) - \theta(z^{N+1})) \tag{3.21}$$

$$\leq \frac{1}{\mu}(\theta(z^0) - \underline{\theta}). \tag{3.22}$$

Taking the limit as $N \to +\infty$ yields the desired summability property. ∎

Before presenting the *inexact optimality* property for any sequence generated by the proposed method, we recall an important result regarding function θ appearing in (3.4), under Assumption 1:

PROPERTY 3.8. *For function θ defined as in (3.4) and satisfying Assumption 1, the following equality holds: for every $(x, p, \beta) \in (\mathbb{R}^n)^3$,*

$$\begin{aligned} &\partial\theta(x, p, \beta) \\ &= \partial_x\theta(x, p, \beta) \times \partial_p\theta(x, p, \beta) \times \partial_\beta\theta(x, p, \beta) \\ &= \{\nabla_x q(x, p, \beta) + \nabla f(x)\} \times \big(\nabla_p q(x, p, \beta) + \partial g(p)\big) \times \big(\nabla_\beta q(x, p, \beta) + \partial h(\beta)\big). \end{aligned}$$

LEMMA 3.9 (Inexact optimality). *Assume that the sequence $(z^{\ell})_{\ell\in\mathbb{N}}$ generated by Algorithm 3.1 is bounded. Then, for every $\ell \in \mathbb{N}$, there exists $b^{\ell} \in \partial\theta(z^{\ell})$ such that*

$$\|b^{\ell}\| \leq \rho\|z^{\ell-1} - z^{\ell}\|, \tag{3.23}$$

where $\rho \in (0, +\infty)$.

Proof. The assumed boundedness obviously implies that there exists a bounded subset S of $\mathbb{R}^n$ such that $\{z^{\ell}\}_{\ell\in\mathbb{N}} = \{(x^{\ell}, p^{\ell}, \beta^{\ell})\}_{\ell\in\mathbb{N}}$, $\{(x^{\ell}, p^{\ell-1}, \beta^{\ell-1})\}_{\ell\in\mathbb{N}}$, and $\{(x^{\ell}, p^{\ell}, \beta^{\ell-1})\}_{\ell\in\mathbb{N}}$ are included in S. In addition, recall that, according to Assumption 1, the coupling term q has a Lipschitz continuous gradient on S. In the following, we will exploit the fact that, at every iteration ℓ, the update for each block of coordinate needs to satisfy Fermat's rule for the corresponding subproblem.

- Fermat's rule for (3.6) reads

$$\gamma_1^{-1} A(x^{\ell-1} - x^{\ell}) = \nabla f(x^{\ell-1}) + \nabla_x q(x^{\ell}, p^{\ell-1}, \beta^{\ell-1}). \tag{3.24}$$

  Notice that

$$\partial_x\theta(x^{\ell}, p^{\ell}, \beta^{\ell}) = \{\nabla_x\theta(x^{\ell}, p^{\ell}, \beta^{\ell})\} = \{\nabla f(x^{\ell}) + \nabla_x q(x^{\ell}, p^{\ell}, \beta^{\ell})\}. \tag{3.25}$$

  So, by defining

$$\begin{aligned} b_x^{\ell} = \gamma_1^{-1} A(x^{\ell-1} - x^{\ell}) + \nabla f(x^{\ell}) + \nabla_x q(x^{\ell}, p^{\ell}, \beta^{\ell}) \\ - \nabla f(x^{\ell-1}) - \nabla_x q(x^{\ell}, p^{\ell-1}, \beta^{\ell-1}), \end{aligned} \tag{3.26}$$

we have $b_x^\ell \in \partial_x \theta(x^\ell, p^\ell, \beta^\ell)$ and

$$\begin{aligned}
&\|b_x^\ell\| \\
&\quad \leq \gamma_1^{-1}\bar{\nu}\|x^{\ell-1} - x^\ell\| + L_f\|x^{\ell-1} - x^\ell\| + L_{q,S}\|(0, p^{\ell-1} - p^\ell, \beta^{\ell-1} - \beta^\ell)\| \\
&\quad = (\gamma_1^{-1}\bar{\nu} + L_f)\|x^{\ell-1} - x^\ell\| + L_{q,S}\|(0, p^{\ell-1} - p^\ell, \beta^{\ell-1} - \beta^\ell)\| \\
&\quad \leq (\gamma_1^{-1}\bar{\nu} + L_f + L_{q,S})\|z^{\ell-1} - z^\ell\|, \qquad (3.27)
\end{aligned}$$

where the first inequality follows from Assumptions 1(i)-(ii) and 2(ii).

- Fermat's rule for (3.7) reads

$$r^\ell + \nabla_p q(x^\ell, p^\ell, \beta^{\ell-1}) + \gamma_2^{-1}(p^\ell - p^{\ell-1}) = 0 \qquad (3.28)$$

where $r^\ell \in \partial g(p^\ell)$. Since

$$r^\ell + \nabla_p q(x^\ell, p^\ell, \beta^\ell) \in \partial_p \theta(x^\ell, p^\ell, \beta^\ell), \qquad (3.29)$$

by defining

$$b_p^\ell = \gamma_2^{-1}(p^{\ell-1} - p^\ell) + \nabla_p q(x^\ell, p^\ell, \beta^\ell) - \nabla_p q(x^\ell, p^\ell, \beta^{\ell-1}), \qquad (3.30)$$

we have $b_p^\ell \in \partial_p \theta(x^\ell, p^\ell, \beta^\ell)$ and

$$\begin{aligned}
\|b_p^\ell\| &\leq \gamma_2^{-1}\|p^{\ell-1} - p^\ell\| + L_{q,S}\|(0, 0, \beta^\ell - \beta^{\ell-1})\| \\
&\leq (\gamma_2^{-1} + L_{q,S})\|z^{\ell-1} - z^\ell\| \qquad (3.31)
\end{aligned}$$

where the first inequality stems from Assumption 1(i).

- Fermat's rule for (3.8) reads

$$s^\ell + \nabla_\beta q(x^\ell, p^\ell, \beta^\ell) + \gamma_3^{-1}(\beta^\ell - \beta^{\ell-1}) = 0 \qquad (3.32)$$

where $s^\ell \in \partial h(\beta^\ell)$. By noticing that

$$s^\ell + \nabla_\beta q(x^\ell, p^\ell, \beta^\ell) \in \partial_\beta \theta(x^\ell, p^\ell, \beta^\ell) \qquad (3.33)$$

and defining

$$b_\beta^\ell = \gamma_3^{-1}(\beta^{\ell-1} - \beta^\ell) \in \partial_\beta \theta(x^\ell, p^\ell, \beta^\ell), \qquad (3.34)$$

we have

$$\|b_\beta^\ell\| \leq \gamma_3^{-1}\|\beta^{\ell-1} - \beta^\ell\| \leq \gamma_3^{-1}\|z^{\ell-1} - z^\ell\|. \qquad (3.35)$$

In a nutshell, by virtue of Property 3.8, $b^\ell = (b_x^\ell, b_p^\ell, b_\beta^\ell) \in \partial\theta(x^\ell, p^\ell, \beta^\ell)$. To conclude, we set

$$\rho = \max\{\gamma_1^{-1}\bar{\nu} + L_f + L_{q,S},\ \gamma_2^{-1} + L_{q,S},\ \gamma_3^{-1}\},$$

which yields the desired inequality (3.23). ∎

We now report a first convergence result for a sequence generated by the proposed algorithm, which is reminiscent from [5, Proposition 6]:

PROPOSITION 3.10 (Properties of the cluster points set). *Suppose that Assumptions 1 and 2 hold. Let $(z^\ell)_{\ell\in\mathbb{N}}$ be a sequence generated by Algorithm 3.1. Denote by $\omega(z^0)$ the (possibly empty) set of its cluster points. Then*

i) if $(z^\ell)_{\ell\in\mathbb{N}}$ is bounded, then $\omega(z^0)$ is a nonempty compact connected set and

$$\operatorname{dist}(z^\ell, \omega(z^0)) \to 0 \quad \text{as} \quad \ell \to +\infty;$$

ii) $\omega(z^0) \subset \operatorname{crit}\theta$, where $\operatorname{crit}\theta$ is the set of critical points of function θ;

iii) θ is finite valued and constant on $\omega(z^0)$, and it is equal to

$$\inf_{\ell\in\mathbb{N}} \theta(z^\ell) = \lim_{\ell\to+\infty} \theta(z^\ell).$$

Proof. The proof of the above results for the proposed algorithm is basically identical to the one for [5, Proposition 6] for PAM algorithm. The only point to check is that our objective function is continuous with respect to x, i.e. the only block of variables on which we apply a different update than in PAM. ∎

In conclusion, we have proved that a bounded sequence generated by the proposed method under Assumptions 1 and 2 satisfies the assumptions in [6, Theorem 2.9]. Hence we can state the following result:

PROPOSITION 3.11. *Let Assumptions 1 and 2 be satisfied and let $(z^\ell)_{\ell\in\mathbb{N}} = \big((x^\ell, p^\ell, \beta^\ell)\big)_{\ell\in\mathbb{N}}$ be a sequence generated by Algorithm 3.1 that is assumed to be bounded. Then*

i) $\sum_{\ell=1}^{+\infty} \|z^{\ell+1} - z^\ell\| < +\infty$;

ii) $(z^\ell)_{\ell\in\mathbb{N}}$ converges to a critical point z^* of θ.

Remark 3.12. It is worth mentioning that the proposed P-SASL-PAM algorithm can be easily adapted to the more general setting of minimizing

$$(\forall X \in \mathbb{R}^N) \quad \theta(X) = q(X) + \sum_{k=1}^{S} g_k(X_k). \tag{3.36}$$

Hereabove, $X = (X_1, \ldots, X_S) \in \mathbb{R}^N$, with each $X_k \in \mathbb{R}^{n_k}$, $k \in \{1, \ldots, S\}$, so that $N = \sum_{k=1}^S n_k$. Function θ involves a a locally Lipschitz-differentiable coupling term $q\colon \mathbb{R}^N \to \mathbb{R}$ and S block-separable terms $g_k\colon \mathbb{R}^{n_k} \to]-\infty, +\infty[$ (with $k \in \{1, \ldots, S\}$), some of which may be differentiable with a Lipschitz continuous gradient. Then, the generalized variant of P-SASL-PAM generates a sequence $(Z^\ell)_{\ell\in\mathbb{N}} = ((X_1^\ell, \ldots, X_S^\ell))_{\ell\in\mathbb{N}}$ where the blocks of coordinates are updated via the following scheme, at every iteration $\ell \in \mathbb{N}$:

For $k = 1, \ldots, S$

$$\begin{cases} X_k^{\ell+1} \in \operatorname{prox}^{A_k}_{\gamma_k q(X_1^{\ell+1}, \ldots, X_{k-1}^{\ell+1}, \cdot, X_{k+1}^{\ell}, \ldots, X_S^{\ell})}(X_k^\ell - \gamma_k A_k^{-1} \nabla g_k(X_k^\ell)) \\ \qquad \text{with } \gamma_k \in (0,1), \text{ if } g_k \text{ is differentiable,} \\ X_k^{\ell+1} \in \operatorname{prox}_{\gamma_k \theta(X_1^{\ell+1}, \ldots, X_{k-1}^{\ell+1}, \cdot, X_{k+1}^{\ell}, \ldots, X_S^{\ell})}(X_k^\ell) \\ \qquad \text{with } \gamma_k > 0, \text{ otherwise.} \end{cases}$$

If, for every $k \in \{1, \ldots, S\}$ such that g_k is Lipschitz differentiable, $A_k \in \mathcal{S}_{n_k}$ satisfies a *majorisation condition* like in Assumption, 2 for function g_k, then the *Sufficient*

Decrease and the *Inexact Optimality* properties expressed in Lemma 3.7 and Lemma 3.9 can be extended to a bounded sequence $(Z^\ell)_{\ell\in\mathbb{N}}$ generated by the above extended variant of P-SASL-PAM. In addition, if function θ is a KŁ function, then the convergence results expressed in Proposition 3.11 can be extended, ensuring that the above mentioned sequence has a finite length, *i.e.* $\sum_{\ell=1}^{\infty}\|Z^{\ell+1}-Z^\ell\|<+\infty$, and it converges to a critical point Z^* of θ.

4. Application of P-SASL-PAM to the Joint Reconstruction and Feature Extraction Problem.

4.1. Smoothing of the coupling term. The application of Algorithm 3.1 to Problem (2.12) requires the involved functions to fulfill the requirements listed in Assumption 1. This section is devoted to this analysis, by first defining the following functions, for every $x=(x_i)_{1\le i\le n}\in\mathbb{R}^n$, $p=(p_i)_{1\le i\le n}\in\mathbb{R}^n$, and $\beta=(\beta_i)_{1\le i\le n}\in\mathbb{R}^n$,

$$\tilde{q}(x,p,\beta)=\sum_{i=1}^{n}|x_i|^{p_i}e^{-\beta_i p_i}, \tag{4.1}$$

$$f(x)=\frac{1}{2\sigma^2}\|y-Kx\|_2^2, \tag{4.2}$$

$$g(p)=\sum_{i=1}^{n}\left(\ln\Gamma(1+\frac{1}{p_i})+\iota_{[a,b]}(p_i)\right)+\lambda\,\mathrm{TV}(p), \tag{4.3}$$

$$h(\beta)=\sum_{i=1}^{n}\left(\beta_i+\frac{\beta_i^2}{2\sigma_\beta^2}\right)+\zeta\,\mathrm{TV}(\beta). \tag{4.4}$$

The first item in Assumption 1 regarding the regularity of the coupling term is not satisfied by (4.1). To circumvent this difficulty, we introduce the *pseudo-Huber loss function* [21] depending on a pair of parameters $\delta=(\delta_1,\delta_2)\in(0,+\infty)^2$ such that $\delta_2<\delta_1$:

$$(\forall t\in\mathbb{R})\quad C_\delta(t)=H_{\delta_1}(t)-\delta_2, \tag{4.5}$$

where H_{δ_1} is the *hyperbolic function* defined, for every $t\in\mathbb{R}$, by $H_{\delta_1}(t)=\sqrt{t^2+\delta_1^2}$. Function (4.5) is used as a smooth approximation of the absolute value involved in (4.1). We then replace (4.1) with

$$q(x,p,\beta)=\sum_{i=1}^{n}\left(C_\delta(x_i)\right)^{p_i}e^{-\beta_i p_i}. \tag{4.6}$$

Function C_δ is infinitely differentiable, i.e. its derivatives are continuous for all orders. Thus function (4.6) satisfies Assumption 1.

Function (4.2) is quadratic convex, thus it clearly satisfies Assumption 1(ii). Function (4.3) is a sum of functions that are proper, lower semicontinuous and either nonnegative or bounded from below. The same applies to function (4.4), which is also strongly convex. It results that (4.3) and (4.4) satisfy Assumption 1(iii).

Now, we must show that Θ is a KŁ function. To do so, let us consider the notion of *o-minimal structure* [66], which is a particular family $\mathcal{O}=\{\mathcal{O}_n\}_{n\in\mathbb{N}}$ where each $\mathcal{O}_n$ is a collection of subsets of $\mathbb{R}^n$, satisfying a series of axioms (we refer to [5, Definition 13], for a complete illustration). We present hereafter the definition of *definable set* and *definable function* in an o-minimal structure:

DEFINITION 4.1 (Definable sets and definable functions). *Given an o-minimal structure $\mathcal{O}$, a set $\mathcal{A} \subset \mathbb{R}^n$ such that $\mathcal{A} \in \mathcal{O}_n$ is said to be definable in $\mathcal{O}$. A real extended valued function $f : \mathbb{R} \to (-\infty, +\infty]$ is said to be definable in $\mathcal{O}$ if its graph is a definable subset of $\mathbb{R}^n \times \mathbb{R}$.*

The importance of these concepts in mathematical optimisation is related to the following key result concerning the Kurdyka-Łojasiewicz property [10]:

THEOREM 4.2. *Any proper lower semicontinuous function $f : \mathbb{R}^n \to (-\infty, +\infty]$ which is definable in an o-minimal structure $\mathcal{O}$ has the KŁ property at each point of* $\operatorname{dom} \partial f$.

Let us identify a structure in which all the functions involved in the definition of Θ are definable. This will be sufficient, as definability is a closed property with respect to several operations including finite sum and composition of functions. Before that, we provide a couple of examples of o-minimal structure. The first is represented by the structure of *globally subanalytic sets* $\mathbb{R}_{\text{an}}$ [36], which contains all the sets of the form $\{(u,t) \in [-1,1]^n \times \mathbb{R} \mid f(u) = t\}$ where $f : [-1,1]^n \to \mathbb{R}$ is an analytic function that can be analytically extended on a neighbourhood of $[-1,1]^n$. The second example is the log-exp *structure* $(\mathbb{R}_{\text{an}}, \exp)$ [66, 71], which includes $\mathbb{R}_{\text{an}}$ and the graph of the exponential function. Even though this second structure is a common setting for many optimisation problems, it does not meet the requirements for ours: as shown in [67], $\Gamma^{>0}$ (*i.e.*, the restriction of the Gamma function to $(0, +\infty)$) is not definable on $(\mathbb{R}_{\text{an}}, \exp)$. We thus consider the larger structure $(\mathbb{R}_{\mathcal{G}}, \exp)$, where $\Gamma^{>0}$ has been proved to be definable [68]. $\mathbb{R}_{\mathcal{G}}$ is an o-minimal structure that extends $\mathbb{R}_{\text{an}}$ and is generated by the class $\mathcal{G}$ of *Gevrey* functions from [64].

We end this section by the following result which will be useful subsequently.

PROPOSITION 4.3. *The function $t \mapsto \ln\Gamma(1+\frac{1}{t})$ defined on $(0, +\infty)$ is μ-weakly convex with $\mu > \mu_0 \approx 0.1136$.*

Proof. Let us show that there exists $\mu > 0$ such that function $t \mapsto \ln\Gamma(1+\frac{1}{t}) + \mu t^2/2$ is convex on $(0, +\infty)$. The second-order derivative of this function on the positive real axis reads

$$\begin{aligned} &\frac{d^2}{dt^2}\left(\ln\Gamma\left(1+\frac{1}{t}\right) + \frac{\mu}{2}t^2\right) \\ &\qquad = \frac{1}{t^3}\left(2\text{Digamma}\left(1+\frac{1}{t}\right) + \frac{1}{t}\text{Digamma}'\left(1+\frac{1}{t}\right) + \mu t^3\right), \end{aligned} \tag{4.7}$$

where the Digamma function is the logarithmic derivative of the Gamma function. In order to show the convexity of the considered function, we need to ensure that (4.7) is positive for every $t \in (0, +\infty)$. By virtue of Bohr–Möllerup's theorem [4, Theorem 2.1], among all functions extending the factorial functions to the positive real numbers, only the Gamma function is log-convex. More precisely, its natural logarithm is (strictly) convex on the positive real axis. This implies that $t \mapsto \text{Digamma}'(t)$ is positive. It results that the only sign-changing term in (4.7) is function $t \mapsto 2\,\text{Digamma}\left(1+\frac{1}{t}\right)$ as $t \mapsto \text{Digamma}(t)$ vanishes in a point $t_0 > 1$ ($t_0 \approx 1.46163$) which corresponds to the minimum point of the Gamma function – and therefore also of its natural logarithm [72]. As a consequence, the Digamma function is strictly positive for $t \in (t_0, +\infty)$, implying that $t \mapsto \text{Digamma}\left(1+\frac{1}{t}\right)$ is strictly positive for all $t \in (0, \frac{1}{t_0 - 1})$. Furthermore, $t \mapsto \text{Digamma}\left(1+\frac{1}{t}\right)$ is strictly decreasing

and bounded from below, as shown by the negativity of its first derivative

$$\frac{d}{dt}\text{Digamma}\left(1+\frac{1}{t}\right) = -\frac{1}{t^2}\text{Digamma}'\left(1+\frac{1}{t}\right)$$

and by the following limit

$$\lim_{t\to+\infty} \text{Digamma}\left(1+\frac{1}{t}\right) = \text{Digamma}(1) = -\mathcal{E}$$

where the last equality holds by virtue of the Gauss Digamma theorem and $\mathcal{E}$ is Euler-Mascheroni's constant $\mathcal{E} \approx 0.57721$ [3]. In conclusion, for $t \in [\frac{1}{t_0-1}, +\infty)$ we need to ensure that the positive terms in (4.7) manage to balance the negative contribution of function $t \mapsto 2\text{Digamma}\left(1+\frac{1}{t}\right) > -2\mathcal{E}$. This leads to a condition on parameter $\mu > 0$, since we can impose that

$$0 < \mu t^3 - 2\mathcal{E},$$

where the right hand side expression has a lower bound $\mu/(t_0-1)^3 - 2\mathcal{E}$ that is positive when

$$\mu > 2\mathcal{E}(t_0-1)^3 = \mu_0 \approx 0.1136.$$

This shows that function $t \mapsto \ln\Gamma(1+\frac{1}{t})$ is μ-weakly convex. ∎

4.2. Proximal computations. Let us now focus on the proximal computations involved in Algorithm 3.1. Given the elaborate structure of the involved functions, no trivial closed-form expression is available to compute the required proximity operators. Luckily, efficient minimisation strategies can be designed to tackle the three inner optimisation problems. To ease the description, we summarize in Algorithm 4.1 the application of Algorithm 3.1 to the resolution of (2.12).

Algorithm 4.1 P-SASL-PAM to solve (2.12)

Initialize x^0, p^0 and β^0
Set $A \in \mathcal{S}_n$
Set $\gamma_1 \in (0,1)$, $\gamma_2 \in (0, 1/\mu_0)$, $\gamma_3 > 0$
For $\ell = 0, 1, \ldots$

$$x^{\ell+1} \in \text{prox}^{A}_{\gamma_1 q(\cdot,p^\ell,\beta^\ell)}(x^\ell - \gamma_1 A^{-1}\nabla f(x^\ell)) \quad \text{(with Alg. 4.2)} \tag{4.8}$$

$$p^{\ell+1} \in \text{prox}_{\gamma_2\theta(x^{\ell+1},\cdot,\beta^\ell)}(p^\ell) \quad \text{(with Alg. 4.4)} \tag{4.9}$$

$$\beta^{\ell+1} \in \text{prox}_{\gamma_3\theta(x^{\ell+1},p^{\ell+1},\cdot)}(\beta^\ell) \quad \text{(with Alg. 4.5)} \tag{4.10}$$

Proximal computation with respect to x. Subproblem (3.6) in Algorithm 3.1 requires the computation of the proximity operator of the following separable function

$$q(\cdot, p^\ell, \beta^\ell) : x \mapsto \sum_{i=1}^{n} (C_\delta(x_i))^{p_i^\ell}\, e^{-\beta_i^\ell p_i^\ell},$$

within a weighted Euclidean metric induced by some matrix $A \in \mathcal{S}_n$. We notice that $x_i \mapsto (C_\delta(x_i))^{p_i^\ell}$ is nonconvex whenever $p_i^\ell \in (0,1)$, for some $i \in \{1,\ldots,n\}$. In

order to overcome this issue, we apply a majorisation principle [62]. Let us introduce function σ defined, for every $u \in [\delta_1, +\infty)$, as $\sigma(u) = (u - \delta_2)^p$ with $p \in (0, 1]$, and vector $\delta = (\delta_1, \delta_2) \in (0, +\infty)^2$ such that $\delta_2 < \delta_1$. Since this function is concave, it can be majorised by its first-order expansion around any point $w > \delta_2$:

$$\begin{aligned} (\forall u > \delta_2) \quad (u - \delta_2)^p &\leq (w - \delta_2)^p + p(w - \delta_2)^{p-1}(u - w), \\ &= (1 - p)(w - \delta_2)^p + p(w - \delta_2)^{p-1}(u - \delta_2). \end{aligned} \tag{4.11}$$

Setting, for every $(t, t') \in \mathbb{R}^2$, $u = H_{\delta_1}(t) \geq \delta_1$, $w = H_{\delta_1}(t') \geq \delta_1$ allows us to deduce the following majorisation:

$$(C_\delta(t))^p \leq (1 - p)(C_\delta(t'))^p + p(C_\delta(t'))^{p-1} C_\delta(t). \tag{4.12}$$

Let us now define $\mathcal{I}^\ell = \{i \in \{1, \ldots, n\} \mid p_i^\ell \geq 1\}$ and $\mathcal{J}^\ell = \{1, \ldots, n\} \setminus \mathcal{I}^\ell$. Given $v = (v_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, we deduce from (4.12) that

$$\begin{aligned} (\forall x = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad q(x, p^\ell, \beta^\ell) &= \sum_{i \in \mathcal{I}^\ell} (C_\delta(x_i))^{p_i^\ell} \, e^{-\beta_i^\ell p_i^\ell} + \sum_{i \in \mathcal{J}^\ell} (C_\delta(x_i))^{p_i^\ell} \, e^{-\beta_i^\ell p_i^\ell} \\ &\leq \overline{q}(x, v, p^\ell, \beta^\ell), \end{aligned} \tag{4.13}$$

where the resulting majorant function is separable, *i.e.*

$$\overline{q}(x, v, p^\ell, \beta^\ell), = \sum_{i=1}^{n} \overline{q}_i(x_i, v_i, p_i^\ell, \beta_i^\ell), \tag{4.14}$$

with, for every $i \in \{1, \ldots, n\}$ and $x_i \in \mathbb{R}$,

$$\begin{aligned} &\overline{q}_i(x_i, v_i, p_i^\ell, \beta_i^\ell) \\ &\quad = \begin{cases} e^{-\beta_i^\ell p_i^\ell} \, (C_\delta(u_i))^{p_i^\ell}, & \text{if } p_i^\ell \geq 1 \\ e^{-\beta_i^\ell p_i^\ell} \Big((C_\delta(v_i))^{p_i^\ell} (1 - p_i^\ell) + p_i^\ell (C_\delta(v_i))^{p_i^\ell - 1} C_\delta(x_i) \Big) & \text{otherwise.} \end{cases} \end{aligned} \tag{4.15}$$

In a nutshell, each term of index $i \in \{1, \ldots, n\}$ in (4.14) coincides either with the i-th term of $q(\cdot, p^\ell, \beta^\ell)$ when $i \in \mathcal{I}^\ell$, or it is a convex majorant of this i-th term with respect to v_i when $i \in \mathcal{J}^\ell$. We thus propose to adopt a majorisation-minimisation procedure by building a sequence of convex surrogate problems for the nonconvex minimisation problem involved in the computation of $\operatorname{prox}^A_{\gamma_1 q(\cdot, p^\ell, \beta^\ell)}$. At the κ-th iteration of this procedure, following the MM principle, the next iterate $x^{\kappa+1}$ is determined by setting $v = x^\kappa$. We summarise the strategy in Algorithm 4.2.

Algorithm 4.2 MM algorithm to approximate $\operatorname{prox}^A_{\gamma_1 q(\cdot, p^\ell, \beta^\ell)}(x^+)$ with $x^+ \in \mathbb{R}^n$

Initialize $x^0 \in \mathbb{R}^n$
For $\kappa = 0, 1, \ldots$ **until** convergence

$$x^{\kappa+1} = \operatorname{prox}^A_{\gamma_1 \overline{q}(\cdot, x^\kappa, p^\ell, \beta^\ell)}(x^+) \quad \text{(with Alg. 4.3)}. \tag{4.16}$$

Function $\overline{q}(\cdot, v, p^\ell, \beta^\ell)$ being convex, proper, and lsc, its proximity operator in the weighted Euclidean metric induced by matrix A is guaranteed to be uniquely defined.

It can be computed efficiently using the Dual Forward-Backward (DFB) method [28], outlined in Algorithm 4.3.

Algorithm 4.3 DFB algorithm to compute $\mathrm{prox}^{A}_{\gamma_1 \overline{q}(\cdot,v;p^\ell,\beta^\ell)}(x^+)$ with $x^+ \in \mathbb{R}^n$

Initialize dual variable $w^0 \in \mathbb{R}^n$
Set $\eta \in (0, 2|||A|||^{-1})$
For $\kappa' = 0, 1, \ldots$ **until** convergence

$$u^{\kappa'} = x^+ - Aw^{\kappa'}, \tag{4.17}$$

$$w^{\kappa'+1} = w^{\kappa'} + \eta u^{\kappa'} - \eta\, \mathrm{prox}_{\eta^{-1}\gamma_1 \overline{q}(\cdot,v,p^\ell,\beta^\ell)}(\eta^{-1} w^{\kappa'} + u^{\kappa'}). \tag{4.18}$$

Return $u^{\kappa'} \in \mathbb{R}^n$

The update in (4.18) can be performed componentwise since function $\overline{q}(\cdot, v, p^\ell, \beta^\ell)$ is separable. Thanks to the separability property, computing $\mathrm{prox}_{\eta^{-1}\gamma_1 \overline{q}(\cdot,v,p^\ell,\beta^\ell)}$ boils down to solve n one-dimensional optimization problems, that is

(4.19) $(\forall u^+ = (u_i^+)_{1\leq i\leq n} \in \mathbb{R}^n)$

$$\mathrm{prox}_{\eta^{-1}\gamma_1 \overline{q}(\cdot,v,p^\ell,\beta^\ell)}(u^+) = \Big(\mathrm{prox}_{\eta^{-1}\gamma_1 \overline{q}_i(\cdot,v_i,p_i^\ell,\beta_i^\ell)}(u_i^+)\Big)_{1\leq i\leq n}.$$

More precisely,

- for every $i \in \{1, \ldots, n\}$, such that $p_i^\ell \leq 1$,

$$\begin{aligned}\mathrm{prox}_{\eta^{-1}\gamma_1 \overline{q}_i(\cdot,v_i,p_i^\ell,\beta_i^\ell)}(u_i^+) &= \mathrm{prox}_{\eta^{-1}\gamma_1 e^{-\beta_i^\ell p_i^\ell} p_i^\ell (C_\delta(v_i))^{p_i^\ell - 1} C_{\delta_1}}(u_i^+) \\ &= \mathrm{prox}_{\eta^{-1}\gamma_1 e^{-\beta_i^\ell p_i^\ell} p_i^\ell (C_\delta(v_i))^{p_i^\ell - 1} H_{\delta_1}}(u_i^+). \end{aligned} \tag{4.20}$$

  The proximity operator of the so-scaled version of function H_{δ_1} can be determined by solving a quartic polynomial equation.[1]
- For every $i \in \{1, \ldots, n\}$ such that $p_i^\ell > 1$,

$$\mathrm{prox}_{\eta^{-1}\gamma_1 \overline{q}_i(\cdot,v_i,p_i^\ell,\beta_i^\ell)}(u_i^+) = \mathrm{prox}_{\eta^{-1}\gamma_1 e^{-\beta_i^\ell p_i^\ell} (C_\delta)^{p_i^\ell}}(u_i^+). \tag{4.21}$$

  The latter quantity can be evaluated through a bisection search to find the root of the derivative of the involved proximally regularised function.

Remark 4.4. Due to the nonconvexity of $q(\cdot, p^\ell, \beta^\ell)$, there is no guarantee that the point estimated by Algorithm 4.3 coincides with the exact proximity point. However, we did not notice any numerical issue in our implementation.

Proximal computation with respect to p. Subproblem (3.7) requires to compute the proximity operator of $\gamma_2\big(q(x^{\ell+1}, \cdot, \beta^\ell) + g\big)$, which is equivalent to solve the following minimization problem

$$\underset{p\in[a,b]^n}{\text{minimize}} \;\; \psi^\ell(p) + \lambda \ell_{1,2}(Dp), \tag{4.22}$$

[1] http://proximity-operator.net/scalarfunctions.html

where, for every $p \in \mathbb{R}^n$, $\psi^\ell(p) = \sum_{i=1}^n \psi_i^\ell(p_i)$ with

$$(\forall i \in \{1,\dots,n\})(\forall p_i \in \mathbb{R}) \tag{4.23}$$
$$\psi_i^\ell(p_i) = \begin{cases} \left(C_\delta(x_i^{\ell+1})\right)^{p_i} e^{-\beta_i^\ell p_i} + \ln\Gamma(1+\frac{1}{p_i}) + \frac{1}{2\gamma_2}(p_i - p_i^\ell)^2 & \text{if } p_i > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, $D = [D_{\mathrm{h}}, D_{\mathrm{v}}]$ where $(D_{\mathrm{h}}, D_{\mathrm{v}}) \in (\mathbb{R}^{n\times n})^2$ are the discrete horizontal and vertical 2D gradient operators, and the $\ell_{1,2}$-norm is defined as

$$(\forall p \in \mathbb{R}^n) \quad \ell_{1,2}(Dp) = \sum_{i=1}^n \|([D_{\mathrm{h}}p]_i, [D_{\mathrm{v}}p]_i)\|_2.$$

Problem (4.22) is equivalent to minimizing the sum of the indicator function of a hypercube, a separable component, and a nonseparable term involving the linear operator D. Accoding to Proposition 4.3, we can ensure the convexity of each term $(\psi_i^\ell)_{1\le i\le n}$ by setting $\gamma_2 < \frac{1}{\mu_0} \approx 8.805$. In order to solve (4.22), it is then possible to implement a Primal-Dual (PD) algorithm [29, 44, 69] as outlined in Algorithm 4.4.

Algorithm 4.4 Primal Dual Algorithm for solving (4.22)

Initialise the dual variables $v_1^0 \in \mathbb{R}^{n\times 2}$,$v_2^0 \in \mathbb{R}^n$.
Set $\tau > 0$ and $\sigma > 0$ such that $\tau\sigma(|||D|||^2 + 1) < 1$.
for $\kappa = 0, 1, \dots$ **until** convergence

$$u^\kappa = p^\kappa - \tau(D^* v_1^\kappa + v_2^\kappa), \tag{4.24}$$
$$p^{\kappa+1} = \mathrm{proj}_{[a,b]^n}(u^\kappa), \tag{4.25}$$
$$w_1^\kappa = v_1^\kappa + \sigma D(2p^{\kappa+1} - p^\kappa), \tag{4.26}$$
$$v_1^{\kappa+1} = w_1^{\kappa+1} - \sigma \mathrm{prox}_{\frac{\lambda\ell_{1,2}}{\sigma}}(\frac{w_1^\kappa}{\sigma}). \tag{4.27}$$
$$w_2^\kappa = v_2^\kappa + \sigma(2p^{\kappa+1} - p^\kappa), \tag{4.28}$$
$$v_2^{\kappa+1} = w_2^{\kappa+1} - \sigma \mathrm{prox}_{\frac{\psi^\ell}{\sigma}}(\frac{w_2^\kappa}{\sigma}). \tag{4.29}$$

Return $p^{\kappa+1} \in [a,b]^n$

The proximity operator of the involved $\ell_{1,2}$ norm has a closed-form expression. For every $w_1 = ([w_1]_{i,1}, [w_1]_{i,2})_{1\le i\le n} \in \mathbb{R}^{n\times 2}$ and $\lambda > 0$, we have

$$\begin{aligned} \mathrm{prox}_{\lambda\ell_{1,2}}(w_1) &= \left(\mathrm{prox}_{\lambda\|\cdot\|_2}\left(([w_1]_{i,1}, [w_1]_{i,2})\right)\right)_{1\le i\le n} \\ &= \left(([w_1]_{i,1}, [w_1]_{i,2}) - \frac{\lambda([w_1]_{i,1}, [w_1]_{i,2})}{\max\{\lambda, \|([w_1]_{i,1}, [w_1]_{i,2})\|_2\}}\right)_{1\le i\le n}. \end{aligned}$$

The proximal point at $w_2^\kappa/\sigma = ([w_2^\kappa]_i/\sigma)_{1\le i\le n} \in \mathbb{R}^n$ of the separable term ψ^ℓ with respect to a step size $1/\sigma$ can be found by minimizing, for every $i \in \{1,\dots,n\}$, the following smooth function

$$(\forall t \in (0,+\infty)) \quad \mathrm{g}_i(t) = \psi_i^\ell(t) + \frac{\sigma}{2}\Big(t - \frac{[w_2^\kappa]_i}{\sigma}\Big)^2.$$

The update in (4.29) then reads

$$v_2^{\kappa+1} = \left([w_2^{\kappa+1}]_i - \sigma[w_2^{\kappa}]_i^*\right)_{1\leq i\leq n}$$

where, for every $i \in \{1,\ldots,n\}$, $[w_2^{\kappa}]_i^*$ corresponds to the unique zero of the derivative of g_i. This zero is found by applying Newton's method initialised with

$$\bar{w}_i = \left(\max\left\{10^{-3}, \frac{[w_2^{\kappa}]_i}{\sigma}\right\}\right)_{1\leq i\leq n}.$$

Proximal computation with respect to β. Subproblem (3.8) requires the solution of the following minimisation problem:

$$\underset{\beta\in\mathbb{R}^n}{\text{minimize}} \;\; \varphi^{\ell}(\beta) + \zeta\ell_{1,2}(D\beta) \tag{4.30}$$

where D and $\ell_{1,2}$ have been defined previously and

$$(\forall \beta = (\beta_i)_{1\leq i\leq n} \in \mathbb{R}^n) \quad \varphi^{\ell}(\beta) = \sum_{i=1}^{n} \varphi_i^{\ell}(\beta_i)$$

with, for every $i \in \{1,\ldots,n\}$,

$$\varphi_i^{\ell}(\beta_i) = \left(C_\delta(x_i^{\ell+1})\right)^{p_i^{\ell+1}} e^{-\beta_i p_i^{\ell+1}} + \beta_i + \frac{\beta_i^2}{2\sigma_\beta^2} + \frac{1}{2\gamma_3}(\beta_i - \beta_i^{\ell})^2$$

The above problem shares structure similar to the one studied in the previous case since the objective function is the sum of the smooth convex term φ^{ℓ} and the nonsmooth convex one $\zeta\,\mathrm{TV} = \zeta\ell_{1,2}(D\cdot)$, and it can be solved by the primal-dual procedure outlined in Algorithm 4.5.

Algorithm 4.5 Primal Dual Algorithm for minimizing (4.30)

Set $\tau > 0$ and $\sigma > 0$ such that $\tau\sigma|||D|||^2 \leq 1$.
Initialise the dual variable $v^0 \in \mathbb{R}^{n\times 2}$.
for $\kappa = 0, 1, \ldots$ **until** convergence

$$u^{\kappa} = \beta^{\kappa} - \tau(D^* v^{\kappa}), \tag{4.31}$$

$$\beta^{\kappa+1} = \mathrm{prox}_{\tau\varphi^{\ell}}(u^{\kappa}), \tag{4.32}$$

$$w^{\kappa} = v^{\kappa} + \sigma D(2\beta^{\kappa+1} - \beta^{\kappa}), \tag{4.33}$$

$$v^{\kappa+1} = w^{\kappa+1} - \sigma\mathrm{prox}_{\frac{\zeta\ell_{1,2}}{\sigma}}\left(\frac{w^{\kappa}}{\sigma}\right). \tag{4.34}$$

Return $\beta^{\kappa+1} \in \mathbb{R}^n$

At each iteration κ of Algorithm 4.5, the proximity operator of φ^{ℓ} is expressed as

$$(\forall \beta = (\beta_i)_{1\leq i\leq n} \in \mathbb{R}^n) \quad \mathrm{prox}_{\tau\varphi^{\ell}}(\beta) = \left(\mathrm{prox}_{\tau\varphi_i^{\ell}}(\beta_i)\right)_{1\leq i\leq n} \tag{4.35}$$

For every $i \in \{1, \dots, n\}$, $\operatorname{prox}_{\tau\varphi_i^\ell}(\beta_i)$ is the minimizer of function

$$(\forall \beta_i \in \mathbb{R}) \quad \mathsf{h}_i(\beta_i) = \varphi_i^\ell(\beta_i) + \frac{1}{2\tau}(\beta_i - u_i^\kappa)^2. \tag{4.36}$$

The nonlinear equation defining the unique zero of the derivative of h_i admits a closed-form solution that involves the Lambert W-function [31]. Indeed, let us introduce the following notation:

$$a_{1,i} = p_i^{\ell+1} \left(C_\delta(x_i^{\ell+1})\right)^{p_i^{\ell+1}}, \tag{4.37}$$

$$a_2 = \left(\frac{1}{\sigma_\beta^2} + \frac{1}{\gamma_3} + \frac{1}{\tau}\right)^{-1}, \tag{4.38}$$

$$a_{3,i} = 1 - \frac{\beta_i^\ell}{\gamma_3} - \frac{u_i^\kappa}{\tau}. \tag{4.39}$$

Then

$$\begin{aligned} \mathsf{h}_i'(\beta_i) = 0 &\iff -a_{1,i}\exp(-p_i^{\ell+1}\beta_i) + \frac{\beta_i}{a_2} + a_{3,i} = 0 \\ &\iff p_i^{\ell+1}(\beta_i + a_2 a_{3,i})\exp(p_i^{\ell+1}(\beta_i + a_2 a_{3,i})) = p_i^{\ell+1} a_{1,i} a_2 \exp(p_i^{\ell+1} a_2 a_{3,i}) \\ &\iff \beta_i = \frac{1}{p_i^{\ell+1}} W(p_i^{\ell+1} a_{1,i} a_2 \exp(p_i^{\ell+1} a_2 a_{3,i})) - a_2 a_{3,i}, \end{aligned} \tag{4.40}$$

where the last equivalence comes from the fact that the Lambert W-function is single valued on satisfies the following identity for a pair $(X, Y) \in \mathbb{R}^2$:

$$X\exp(X) = Y \iff X = W(Y). \tag{4.41}$$

In conclusion, the update in (4.32) reads as $\beta^{\kappa+1} = \left(\beta_i^{\kappa+1}\right)_{1\le i\le n}$ where each component of this vector is calculated according to (4.40).

5. Numerical Experiments. We now illustrate the performance of P-SASL-PAM by means of two examples of joint segmentation/deblurring of textured images (Sec. 5.1) and ultrasound images (Sec. 5.2). Let us first explain how our approach (common to both examples) is practically implemented in the context of joint segmentation/deblurring.

The observation model reads as (2.1), and the goal is to retrieve an estimate of the sought image as well as a segmented version of it. The standard deviation σ of the noise affecting the data and the linear operator K are assumed to be known. We adopt the recovery strategy described in Section 4. We describe hereafter the setting of the model/algorithm hyperparameters.

The model parameters that need to be tuned are the regularisation parameters $(\lambda, \zeta) \in]0, +\infty[^2$ for the TV terms, the $\delta_1 > 0$ and $\delta_2 > 0$ values for the pseudo Huber function, and the standard deviation $\sigma_\beta > 0$ for the reparameterised scale parameter. Parameters (λ, ζ) are identified via an empirical search based on visual inspection, considering the fact that the higher the effect of the TV regularisation term, the flatter the estimated solution is. The third parameter, $\delta = (\delta_1, \delta_2)$, is tuned so that δ_1 is chosen in the range $[0.1, 3]$ as the one that defines the best trade-off between a high PSNR and a high overall accuracy (OA) value, while $\delta_2 = \delta_1 - 0.01$. Eventually, for the last parameter σ_β, experimental results give credit to the fact that the choice $\sigma_\beta = 1$ is a robust one, so it is used in all our experiments.

The algorithmic hyperparameters include the stepsizes of the proximal steps, as well as the preconditioning matrix involved in the preconditioned proximal gradient step. We set $(\gamma_1, \gamma_2, \gamma_3) = (0.99, 1, 1)$. With this choice of γ_2, the condition $\gamma_2 < 8.805$ for the convexity of the function in (4.22) is satisfied. For the preconditioner, we consider a regularized version of the inverse of the Hessian of the data fidelity function in (4.2), given by

$$A = \sigma^2 (K^\top K + \mu \mathbb{I}_m)^{-1}$$

where $\mu = 0.1$, so that A is well defined.

In order to obtain the labelling of a segmented image from our estimated shape parameter (denoted by $\widehat{p}$) we use a quantisation procedure based on Matlab functions `multithresh` and `imquantize`. The former defines a desired number of quantisation levels using Otsu's method, while the latter performs a truncation of the data values according to the provided quantisation levels. We remark here that the number of labels does not need to be defined throughout the proposed optimisation procedure, but only at the final segmentation step. This step can thus be considered as a post-processing that is performed on the estimated solution.

In order to evaluate the quality of the solution, we consider the following metrics: for the estimated image, we make use of the peak signal-to-noise ratio (PSNR) defined as follows, x being the original signal and $\widehat{x}$ the estimated one:

$$\mathrm{PSNR} = 10 \log_{10} \big(n \max_{i \in \{1,\dots,n\}} (x_i, \hat{x}_i)^2 / \|x_i - \hat{x}_i\|^2 \big),$$

and of the structure similarity measure (SSIM) [70]. For the segmentation task we compute the percentage OA of correctly predicted labels.

The stopping criterions for both the P-SASL-PAM outer loop and the inner loops are set by defining a threshold level on the relative change between two consecutive iterates of the involved variables and a maximum number of iterations. The outer loop in Algorithm 3.1 stops whenever $\ell = 2000$ or when $\|z^{\ell+1} - z^\ell\| / \|z^\ell\| < 10^{-4}$. The MM procedure to compute $x^{\ell+1}$ in Algorithm 4.2 is stopped after 300 iterations or when $\|x^{\kappa+1} - x^\kappa\| / \|x^\kappa\| < 5 \times 10^{-3}$. The DFB procedure in Algorithm 4.3 to compute $u^{\kappa+1}$ is stopped after 300 iterations or when $\|u^{\kappa+1} - u^\kappa\| / \|u^\kappa\| < 10^{-3}$. The PD procedure in Algorithms 4.4 and 4.5 computing $p^{\ell+1}$ (resp. $\beta^{\ell+1}$) terminates after 200 iteration or when $\|p^{\kappa+1} - p^\kappa\| / \|p^\kappa\| < 10^{-3}$ (resp. $\|\beta^{\kappa+1} - \beta^\kappa\| / \|\beta^\kappa\| < 10^{-3}$).

In the first example, we illustrate the performance of the proposed method on an image that is composed of three different texture regions. For the second example, we work on the ultrasound images considered in [30, Section IV.C] and provide a quantitative comparison with the methods that are mentioned in this work - namely, a combination of Wiener deconvolution and Otsu's segmentation [54], a combination of Lasso deconvolution and SLaT segmentation [15], the adjusted Hamiltonian Monte Carlo (HMC) method [60], the Proximal Unadjusted Langevin algorithm (P-ULA) [56] and its preconditioned version (PP-ULA) [30] for joint deconvolution and segmentation.

5.1. Example 1. In this first example, we propose an illustration of the ability of the proposed framework to deconvolve and segment a synthetically created image. The image, which we refer to as *Texture*, is a combination of three textures belonging to the Original Brodatz's Database[2] [1]. Each texture is located in a distinct region,

[2] https://multibandtexture.recherche.usherbrooke.ca/original_brodatz.html

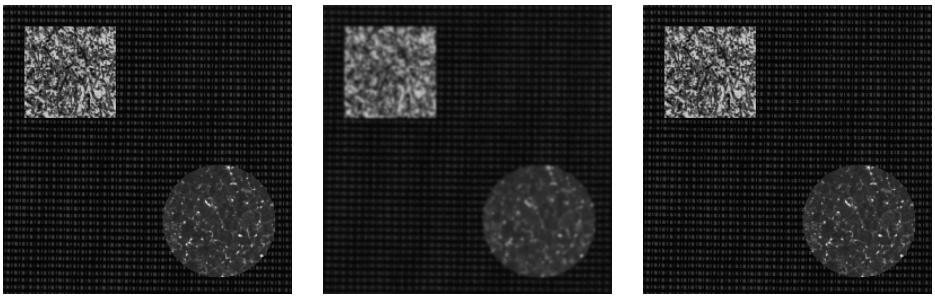

FIG. 1. *Original (x), Degraded (y) and Reconstructed ($\hat{x}$) versions of* Texture.

namely the background, a quadratic shape in the upper left corner and a circular shape in lower right corner. We display the resulting image in Figure 1(left). We assume that each texture is characterised by different $\mathcal{GGD}$ parameters and, in particular, we constrain the shape parameter to the interval $[0.01, 10]$. For the degradation process we choose K as the blur operator corresponding to the convolution with an isotropic Gaussian filter of size 7×7 and standard deviation 1, created with Matlab function `fspecial`. Furthermore, we utterly corrupt the data with additive white Gaussian noise with zero mean and small standard deviation $\sigma = 0.1$. The degraded image is displayed in Figure 1(middle).

As a starting point for the algorithm we choose x^0 as the degraded image y, $(p_i^0)_{1\leq i\leq n}$ is drawn from an i.i.d. uniform random distribution over the range $[0.5, 1.5]$, and $(\beta_i^0)_{1\leq i\leq n}$ is drawn from an i.i.d Gaussian distribution with zero mean and standard deviation $\sigma_\beta = 1$. Figure 1(right) the reconstructed image $\hat{x}$, using P-SASL-PAM. Figure 2 shows the ability of P-SASL-PAM to accurately reconstruct a piecewise constant approximation of the shape parameter p. Namely, we propose a comparison between the reference labelling of the textures in the image (Fig. 2(left)) and the estimated shape parameter $\hat{p}$ (Fig. 2(middle)) along with its quantised version $\bar{p}$ (Fig. 2(right)), which corresponds to our estimated labelling. Figure 3 shows the decay of the cost function θ along the first 500 iterations of P-SASL-PAM, assessing its fast and stable convergence. Eventually, we report in Table 1 the metrics of the quantitative evaluation of the estimated solution, which confirms the good performance of our proposed method.

PSNR	SSIM	OA
33.13	0.95	99.7

TABLE 1
PSNR, SSIM and OA for Texture

5.2. Example 2. In this example, we illustrate the good performance of our approach on the joint deconvolution/segmentation of realistically simulated ultrasound images with two regions (*Simu1*) and three regions (*Simu2*) extracted from [30]. We define K as the linear operator modelling the convolution with the point spread function of the probe, and set the noise variance to $\sigma^2 = 0.013$ for *Simu1* and $\sigma^2 = 33$ for *Simu2*. Following the procedure outlined in [30], we initialise x^0 using

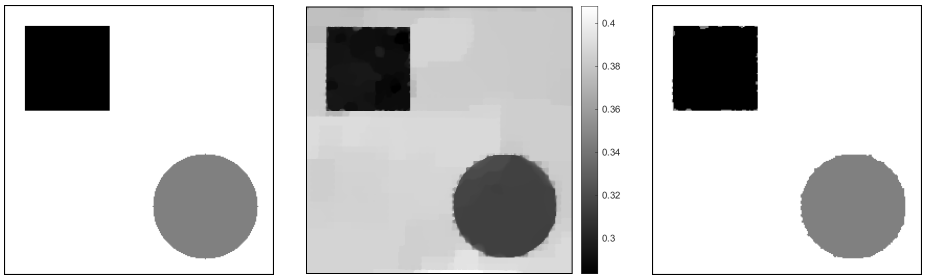

FIG. 2. *Segmentation of the shape parameter for* Texture*: reference labelling, estimated* $\hat{p}$ *and quantised* $\bar{p}$*.*

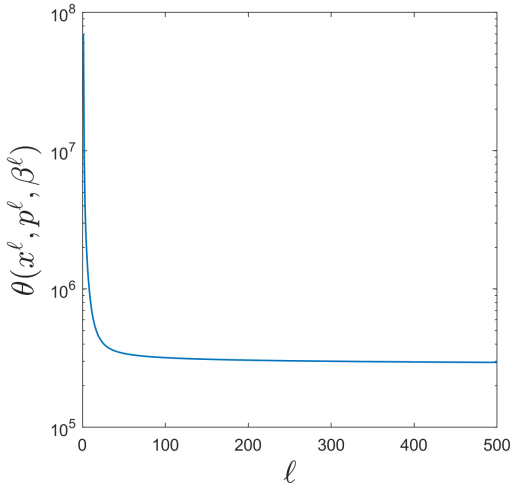

FIG. 3. *Texture: Decay of the objective value along 500 iterations.*

a pre-deconvolved image obtained with a Wiener filter applied to the observed data y, $(p_i^0)_{1\leq i\leq n}$ is drawn from an i.i.d. uniform distribution in the range $[0.5, 1.5]$, while $(\beta_i^0)_{1\leq i\leq n}$ is drawn from an i.i.d. Gaussian distribution with zero mean and unit standard deviation.

Figure 4 illustrates the B-mode image of the original x, of the degraded y, and of the reconstructed image $\hat{x}$ on both examples. The B-mode image is the most common representation of an ultrasound image, displaying the acoustic impedance of a 2-dimensional cross section of the considered tissue. The reconstructed results in Figure 4(right) show clearly reduced blur and sharper region contours. We then report in Figure 5 the segmentation obtained from the estimated shape parameter via the aforementioned quantisation procedure, confirming its great capabilities. Figure 6 shows a plot of the vectorised references for the shape parameter against the vectorised versions of the images obtained by assigning to each estimated region the median of the p_i values within it. We notice that our estimated median values are consistent with the original ones. The results for *Simu2* are slightly less accurate, but this is in agreement with the results presented in [30, Table III] for P-ULA, HMC and PP-ULA, suggesting that the configuration of the parameters for *Simu2* is quite challenging.

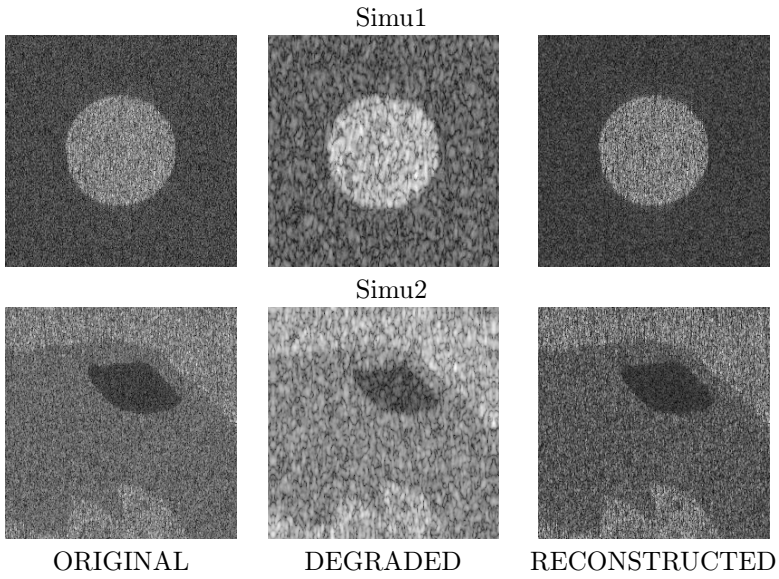

FIG. 4. *B-mode of Simu1 and Simu2. The B-mode image is the most common type of ultrasound image, displaying the acustic impedance of a 2-dimensional cross section of the considered tissue. All images are presented in the same scale [0,1].*

Figure 7 shows the evolution of the cost function for both Simu1 and Simu2 along 1000 iterations, hereagain showing the great convergence behavior of our algorithm.

Eventually, Tables 2 and 3 propose a quantitative comparison of our results against those of the methods considered in [30]. From these tables we can conclude that the proposed variational method is able to compete with state-of-the-art Monte Carlo Markov Chain techniques in terms of both segmentation and deconvolution performance.

METHOD	PSNR	SSIM	OA
Wiener-Otsu	37.1	0.57	99.5
Lasso-SLaT	39.2	0.60	99.6
P-ULA	38.9	0.45	98.7
HMC	40.0	0.62	99.7
PP-ULA	40.3	0.62	99.7
OURS	40.2	0.61	99.9

TABLE 2
PSNR, SSIM and OA scores for Simu1

6. Conclusions. We investigated a novel approach for the joint reconstruction-feature extraction problem. The novelty in this work lies both in the problem formulation and in the resolution procedure. Firstly, we proposed a new variational model in which we introduced a flexible sparse regularisation term for the reconstruction task; secondly, we designed a new iterative block alternating minimization method, whose aim is to exploit the structure of the problem and the properties of the functions involved in it. We established convergence results for the proposed algorithm and illustrated the validity of the approach on numerical examples in the case of

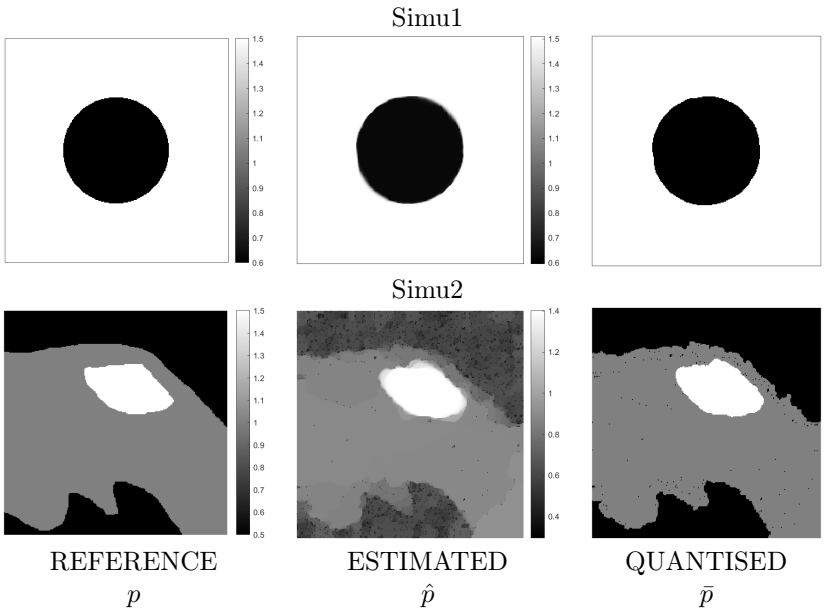

FIG. 5. *Segmentation of the shape parameter for* Simu1 *and* Simu2*: reference p, estimated $\hat{p}$ and quantised $\bar{p}$.*

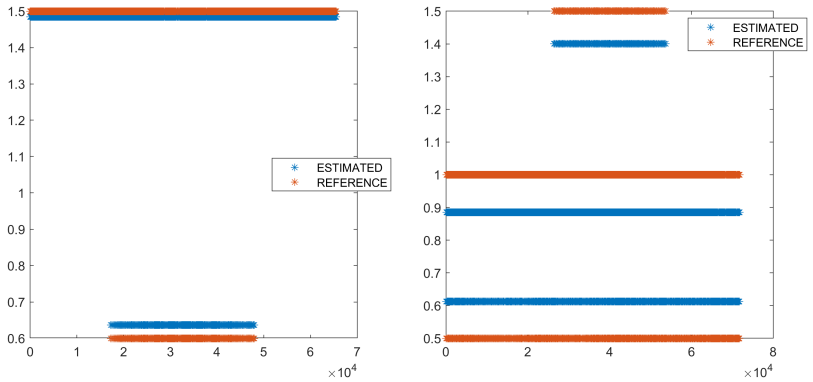

FIG. 6. *Plot of the vectorised estimated shape parameter median values (blue) against the reference values (red) for* Simu1 *(a) and* Simu2 *(b).*

joint deconvolution-segmentation problems. We also included comparisons with state-of-the-art methods with respect to which our proposal registers a similar and even superior qualitative performance. An attractive aspect of the proposed work is that the space variant parameters defining the flexible sparse regularisation do not need to be defined in advance, but are inherently estimated by the iterative optimisation procedure.

Acknowledgments. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861137. The authors thank Ségolène Martin for her careful reading of the initial version of this manuscript.

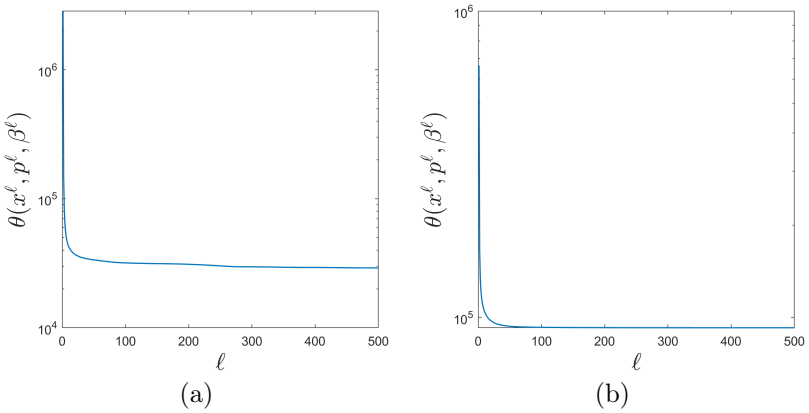

FIG. 7. *Decay of the objective value along 500 iterations for* Simu1 *(a) and* Simu2 *(b).*

METHOD	PSNR	SSIM	OA
Wiener-Otsu	35.4	0.63	96.0
Lasso-SLaT	37.8	0.70	98.3
P-ULA	37.1	0.57	94.9
HMC	36.4	0.64	98.5
PP-ULA	38.6	0.71	98.7
OURS	37.9	0.67	97.5

TABLE 3
PSNR, SSIM and OA scores for Simu2

REFERENCES

[1] S. ABDELMOUNAIME AND H. DONG-CHEN, *New Brodatz-based image databases for grayscale color and multiband texture analysis*, International Scholarly Research Notices, 2013 (2013).
[2] M. AHOOKHOSH, L. T. K. HIEN, N. GILLIS, AND P. PATRINOS, *Multi-block Bregman proximal alternating linearized minimization and its application to orthogonal nonnegative matrix factorization*, Computational Optimization and Applications, 79 (2021), p. 681–715.
[3] G. E. ANDREWS, R. ASKEY, AND R. ROY, *Special Functions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1999.
[4] E. ARTIN, *The Gamma Function*, Athena series, Holt, Rinehart and Winston, 1964.
[5] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), p. 438–457.
[6] H. ATTOUCH, J. BOLTE, AND B. F. SVAITER, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming, Series A, 137 (2011), pp. 91–124.
[7] D. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, 2 ed., 1999.
[8] P. BLOMGREN, T. CHAN, P. MULET, AND C. WONG, *Total variation image restoration: Numerical methods and extensions*, IEEE International Conference on Image Processing, (1997).
[9] J. BOLTE, H. BAUSCHKE, AND M. TEBOULLE, *A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications*, Mathematics of Operations Research, 42 (2016).
[10] J. BOLTE, A. DANIILIDIS, A. LEWIS, AND M. SHIOTA, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization, 18 (2007), pp. 556–572.
[11] J. BOLTE, S. SABACH, AND M. TEBOULLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.
[12] J. BOLTE, S. SABACH, M. TEBOULLE, AND Y. VAISBOURD, *First order methods beyond con-*

vexity and Lipschitz gradient continuity with applications to quadratic inverse problems, SIAM Journal on Optimization, 28 (2018), pp. 2131–2151.

[13] S. Bonetti, M. Prato, and S. Rebegoldi, *A block coordinate variable metric linesearch based proximal gradient method*, Computational Optimization and Applications, (2018).

[14] S. Bonettini, F. Porta, M. Prato, S. Rebegoldi, V. Ruggiero, and L. Zanni, *Recent Advances in Variable Metric First-Order Methods*, Springer International Publishing, Cham, 2019, pp. 1–31.

[15] X. Cai, R. Chan, M. Nikolova, and T. Zeng, *A three-stage approach for segmenting degraded color images: Smoothing, Lifting and Thresholding (SLaT)*, Journal of Scientific Computing, 72 (2017), pp. 1313–1332.

[16] X. Cai, R. Chan, C.-B. Schönlieb, G. Steidl, and T. Zeng, *Linkage between piecewise constant Mumford–Shah model and Rudin–Osher–Fatemi model and its virtue in image segmentation*, SIAM Journal on Scientific Computing, 41 (2019), pp. B1310–B1340.

[17] X. Cai, R. Chan, and T. Zeng, *A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 368–390.

[18] Y. Censor and A. Lent, *Optimization of "log x" entropy over linear equality constraints*, Siam Journal on Control and Optimization, 25 (1987), pp. 921–933.

[19] A. Chambolle, D. Cremers, and T. Pock, *A convex approach to minimal partitions*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 1113–1158.

[20] R. Chan, H. Yang, and T. Zeng, *A two-stage image segmentation method for blurry images with Poisson or multiplicative Gamma noise*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 98–127.

[21] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Deterministic edge-preserving regularization in computed imaging*, IEEE Transactions on Image Processing, 6 2 (1997), pp. 298–311.

[22] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Processing Letters, 14 (2007), pp. 707 – 710.

[23] Y. Chen, S. Levine, and M. Rao, *Variable exponent, linear growth functionals in image restoration*, SIAM Journal on Applied Mathematics, 66 (2006), pp. 1383–1406.

[24] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, *Variable metric Forward-Backward algorithm for minimizing the sum of a differentiable function and a convex function*, Journal of Optimization Theory and Applications, 162 (2014), pp. 107–132.

[25] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, *A block coordinate variable metric forward–backward algorithm*, Journal of Global Optimization, (2016), pp. 1–29.

[26] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, Springer New York, New York, NY, 2011, pp. 185–212.

[27] P. L. Combettes and J.-C. Pesquet, *Fixed point strategies in data science*, IEEE Transactions on Signal Processing, 69 (2021), p. 3878–3905.

[28] P. L. Combettes, Đinh Dũng, and Bằng Công Vũ, *Proximity for sums of composite functions*, Journal of Mathematical Analysis and Applications, 380 (2011), pp. 680–688.

[29] L. Condat, *A Primal–Dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013).

[30] M.-C. Corbineau, D. Kouamé, E. Chouzenoux, J.-Y. Tourneret, and J.-C. Pesquet, *Preconditioned P-ULA for joint deconvolution-segmentation of ultrasound images*, IEEE Signal Processing Letters, 26 (2019), pp. 1456–1460.

[31] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, *On the Lambert W function*, Advances in Computational Mathematics, 5 (1996), pp. 329–359.

[32] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constrains*, Communications on Pure and Applied Mathematics, 57 (2004).

[33] M. Do and M. Vetterli, *Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance*, IEEE Transactions on Image Processing, 11 (2002), pp. 146–158.

[34] H. Erdogan and J. Fessler, *Monotonic algorithms for transmission tomography*, IEEE Transactions on Medical Imaging, 18 (1999), pp. 801–814.

[35] M. Foare, N. Pustelnik, and L. Condat, *Semi-Linearized proximal alternating minimization for a discrete Mumford–Shah model*, IEEE Transactions on Image Processing, 29 (2020), pp. 2176–2189.

[36] A. Gabrielov, *Complements of subanalytic sets and existential formulas for analytic functions*, Inventiones mathematicae, 125 (1996), pp. 1–12.

[37] D. Ghilli and K. Kunisch, *On monotone and primal-dual active set schemes for ℓ_p-type problems, $p \in (0, 1]$*, Computational Optimization and Applications, 72 (2019), pp. 45–85.
[38] M. Grasmair, *Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term*, Inverse Problems & Imaging, 3 (2009), pp. 383–387.
[39] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with lq penalty term*, Inverse Problems, 24 (2008), p. 055020.
[40] J. Hertrich and G. Steidl, *Inertial stochastic palm (ispalm) and applications in machine learning*, (2020).
[41] C. Hildreth, *A quadratic programming procedure*, Naval Research Logistics Quarterly, 4 (1957), pp. 79–85.
[42] M. Hintermüller and T. Wu, *Nonconvex TVq-models in image restoration: analysis and a Trust-Region regularization-based superlinearly convergent solver*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1385–1415.
[43] D. Hunter and K. Lange, *A Tutorial on MM Algorithms*, The American Statistician, 58 (2004), pp. 30–37.
[44] N. Komodakis and J.-C. Pesquet, *Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems*, IEEE Signal Processing Magazine, 32 (2015), pp. 31–54.
[45] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'Institut Fourier, 48 (1998), pp. 769–783.
[46] A. Lanza, S. Morigi, M. Pragliola, and F. Sgallari, *Space-variant generalised Gaussian regularisation for image restoration*, Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 7 (2018), pp. 1–14.
[47] H. Le, N. Gillis, and P. Patrinos, *Inertial block proximal methods for non-convex non-smooth optimization*, 119 (2020), pp. 5671–5681, https://proceedings.mlr.press/v119/le20a.html.
[48] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels.* Equ. Derivees partielles, Paris 1962, Colloques internat. Centre nat. Rech. sci. 117, 87-89 (1963)., 1963.
[49] S. Łojasiewicz, *Sur la géométrie semi- et sous- analytique*, Annales de l'Institut Fourier, 43 (1993), pp. 1575–1595.
[50] D. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, Journal of Inverse and Ill-posed Problems, 16 (2008), pp. 463–478.
[51] D. Lorenz and E. Resmerita, *Flexible sparse regularization*, Inverse Problems, 33 (2016).
[52] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on Pure and Applied Mathematics, 42 (1989), pp. 577–685.
[53] M. Nikolova and P. Tan, *Alternating proximal gradient descent for nonconvex regularised problems with multiconvex coupling terms.* Aug. 2017, https://hal.archives-ouvertes.fr/hal-01492846.
[54] N. Otsu, *A threshold selection method from gray-level histograms*, IEEE Transactions on Systems, Man, and Cybernetics, 9 (1979), pp. 62–66.
[55] B. Pascal, S. Vaiter, N. Pustelnik, and P. Abry, *Automated data-driven selection of the hyperparameters for total-variation based texture segmentation*, Journal of Mathematical Imaging and Vision, 63 (2021), pp. 923–952.
[56] M. Pereyra, *Proximal Markov chain Monte Carlo algorithms*, Statistics and Computing, 26 (2013).
[57] T. Pock and S. Sabach, *Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1756–1787.
[58] R. Ramlau and E. Resmerita, *Convergence rates for regularization with sparsity constraints*, Electronic transactions on numerical analysis ETNA, 37 (2010), pp. 87–104.
[59] A. Repetti and Y. Wiaux, *Variable metric forward-backward algorithm for composite minimization problems*, SIAM Journal on Optimization, 31 (2021), pp. 1215–1241.
[60] C. Robert, V. Elvira, N. Tawn, and C. Wu, *Accelerating MCMC algorithms*, Wiley Interdisciplinary Reviews: Computational Statistics, 10 (2018).
[61] R. Rockafellar and R. J. Wets, *Variational Analysis*, Springer Verlag, 2004.
[62] E. D. Schifano, R. L. Strawderman, and M. T. Wells, *Majorization-Minimization algorithms for nonsmoothly penalized objective functions*, Electronic Journal of Statistics, 4 (2010), pp. 1258 – 1299.
[63] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society: Series B (Methodological), 58 (1996), pp. 267–288.
[64] J. Tougeron, *Sur les ensembles semi-analytiques avec conditions gevrey au bord*, Annales

Scientifiques De L Ecole Normale Superieure, 27 (1994), pp. 173–208.
[65] P. Tseng, *Convergence of a block coordinate descent method for nondifferentiable minimization*, Journal of Optimization Theory and Applications, 109 (2001), pp. 475–494.
[66] L. Van Den Dries, *Tame Topology and O-minimal Structures*, London Mathematical Society Lecture Note Series, Cambridge University Press, 1998.
[67] L. Van Den Dries, A. Macintyre, and D. Marker, *Logarithmic-exponential power series*, Journal of the London Mathematical Society, 56 (1997), pp. 417–434.
[68] L. Van Den Dries and P. Speissegger, *The field of reals with multisummable series and the exponential function*, Proceedings of The London Mathematical Society, 81 (2000), pp. 513–565.
[69] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics, 38 (2013), pp. 667–681.
[70] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612.
[71] A. Wilkie, *Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function*, Journal of the American Mathematical Society, 9 (1996), pp. 1051–1094.
[72] J. W. Wrench, *Concerning two series for the Gamma function*, Mathematics of Computation, 22 (1968), pp. 617–626.
[73] C. Zarzer, *On Tikhonov regularization with non-convex sparsity constraints*, Inverse Problems, 25 (2009), p. 025006.
[74] N. Zhao, A. Basarab, D. Kouamé, and J.-Y. Tourneret, *Joint segmentation and deconvolution of ultrasound images using a hierarchical Bayesian model based on generalized Gaussian priors*, IEEE Transactions on Image Processing, 25 (2016), pp. 3736–3750.