



HAL
open science

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

► To cite this version:

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. 2022. hal-03591742v1

HAL Id: hal-03591742

<https://hal.science/hal-03591742v1>

Preprint submitted on 28 Feb 2022 (v1), last revised 18 Mar 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

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

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

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

17 **AMS subject classifications.**

18 **1. Introduction.** Variational regularisation of ill-posed inverse problems in ima-
19 ging relies on the idea of searching for a solution in a well-suited space. A central
20 role in this context is played by ℓ_p spaces with $p \in (0, \infty)$, and the power p of the
21 corresponding norms when $p \geq 1$ [32, 39, 50, 58, 63] or seminorms when $p \in (0, 1)$
22 [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-
23)norm is denoted by $\|u\|_p = (\sum_{i=1}^n |u_i|^p)^{1/p}$. We usually omit p when $p = 2$, so that
24 $\|\cdot\| = \|\cdot\|_2$. The case $p \in (0, 1)$ has gained rising credit, especially in the field of sparse
25 regularisation. An extensive literature has been focused on challenging numerical
26 tasks raised by the nonconvexity of the seminorms and the possibility to combine
27 them with linear operators to extract salient features of the sought images [37, 42]. In
28 [51] the more general notion of F -norm is introduced in order to establish functional
29 analysis results on products of ℓ_{p_i} -spaces with $p_i \in (0, 2]$. For some $x = (x_i)_{1 \leq i \leq n} \in$
30 \mathbb{R}^n , this amounts to studying the properties of penalties of the form $\sum_{i=1}^n |x_i|^{p_i}$, for
31 some positive exponents $(p_i)_{1 \leq i \leq n}$. This approach offers a more flexible framework
32 by considering a wider range of exponents than the standard ℓ_p -based regularisation.
33 However, it extends the problem of choosing a suitable exponent p to a whole sequence
34 of exponents $(p_i)_{1 \leq i \leq n}$. In image restoration, a related approach consists in adopting
35 space variant regularisation models built around a Total Variation-like functional
36 with a variable exponent $\sum_{i=1}^n \|(\nabla x)_i\|^{p_i}$ where ∇ is a discrete 2D gradient operator.
37 The rationale is to select the set of parameters $(p_i)_{1 \leq i \leq n}$ in order to promote either
38 edge enhancement ($p_i = 1$) or smoothing ($p_i > 1$) depending on the spatial location
39 encoded by index i . This model was introduced in [8] and then put into practice
40 firstly for $p_i \in [1, 2]$ in [23] and then for $p_i \in (0, 2]$ in [46]. In all of these works,
41 the so-called space variant p -map (i.e., $(p_i)_{1 \leq i \leq n}$) is estimated offline in a preliminary
42 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

43 In this paper, we address the problem of joint image recovery and features ex-
 44 traction. Image recovery amounts to retrieving an estimate of an original image from
 45 a degraded version of it. The degradation usually corresponds to the application of
 46 a linear operator (e.g., blur, projection matrix) to the image, and the addition of a
 47 noise. Feature extraction problems arise when one wants to assign to an image a small
 48 set of parameters which can describe or identify the image itself. Image segmentation
 49 can be viewed as an example of features extraction, which consists of defining a label
 50 field on the image domain so that pixels are partitioned into a predefined number of
 51 homogeneous regions according to some specific characteristics. Texture retrieval is a
 52 second example. This task relies on the idea of assigning a set of parameters to each
 53 coefficients of the image, in some transformed space, so that the combination of all
 54 parameters defines a “signature” that represents the content of various spatial regions.
 55 Joint image recovery and feature extraction consists in performing, in a joint manner,
 56 the image recovery and the extraction of features in the sought image. A power-
 57 ful and versatile approach for features extraction, that we will adopt here, assumes
 58 that the image data follow a mixture of generalised Gaussian probability distribution
 59 (\mathcal{GGD}) [30, 33, 74]. In our case, this leads to the minimization of a non-smooth and
 60 non-convex cost function. The \mathcal{GGD} model results in a sum of weighted ℓ_{p_i} -based
 61 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
 62 thus aim at jointly estimating an optimal configuration for $(\vartheta_i, p_i)_{1 \leq i \leq n}$, and retriev-
 63 ing the image. Under an assumption of consistency within the exponents values of
 64 a given region of the features space, we indeed obtain the desired feature extraction
 65 starting from the estimated p -map.

66 The specific structure of the objective function we will propose suggests the use
 67 of an alternating minimisation procedure. In such an approach, one sequentially up-
 68 dates a subset of parameters through the resolution of an inner minimization problem,
 69 while the other parameters are assumed to be fixed. This approach has a standard
 70 form in the Block Coordinate Descent method (BCD) (also known as Gauss-Seidel
 71 algorithm) [41]. In the context of nonsmooth and nonconvex problems, the simple
 72 BCD may show instabilities [65], which resulted in an extensive construction of al-
 73 ternative methods that efficiently exploit the characteristics of the functions, and
 74 introduce powerful tools to improve the convergence guarantees of BCD, or overcome
 75 difficulties arising in some formulations. In this respect, a central role is played
 76 by proximal methods [26, 27]: a proximally regularised BCD (PAM) for nonconvex
 77 problems was studied in [5]; a proximal linearised method (PALM) and its inertial
 78 and stochastic versions were then proposed in [11] resp. [57] and [40]; in [35] the
 79 authors investigated the advantage of a hybrid semi-linearised scheme (SL-PAM) for
 80 the joint task of image restoration and edge detection based on a discrete version of
 81 the Mumford–Shah model. A structure-adapted version of PALM (ASAP) was de-
 82 signed in [53] to exploit the block-convexity of the coupling terms and the regularity
 83 of the block-separable terms arising in some practical applications such as nonneg-
 84 ative matrix factorisation and blind source separation. The extension to proximal
 85 mappings defined w.r.t. a variable metric was firstly introduced in [18], leading to the
 86 so-called Block Coordinate Variable Metric Forward Backward. An Inexact version
 87 and a linesearch based version of it were presented in [25] and [13], respectively. In
 88 [59] the authors introduced a Majorisation-Minimisation strategy in a Variable Metric
 89 Forward-Backward algorithm to tackle the challenging task of computing the prox-
 90 imity operator of composite functions. We refer to [14] for an in-depth analysis of
 91 how to introduce a variable metric into first-order methods. To conclude this brief
 92 overview, let us also mention generalisations to proximal mappings defined according

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

96 In the problem formulation we will be interested in, the objective function includes
 97 a quadratic term, hence Lipschitz differentiable, that is restricted to a single block
 98 of parameters. This feature makes the related subproblem well-suited for a splitting
 99 procedure that involves an explicit gradient step with respect to this term. None of
 100 the aforementioned methods provide a proper framework for this purpose. In order to
 101 exploit the described particular structure, we thus propose a new BCD method that
 102 mixes standard and linearised proximal regularisations on the different blocks (as in
 103 SL-PAM). The novelty lies in the fact that we include a preconditioned and structure-
 104 adapted linearised step to obtain more efficient and faster proximal computations (as
 105 in ASAP). We refer to the proposed method as the Preconditioned Semi-Linearised
 106 Structure Adapted Proximal Alternating Minimisation (P-SASL-PAM). We investi-
 107 gate the convergence properties for this algorithm based on the framework designed in
 108 [6]. Under analytical assumptions on the objective function, we show the global con-
 109 vergence toward a critical point of any sequence generated by the proposed method.
 110 Then, we explicit the use of this method in our problem of image recovery and feature
 111 extraction. The performance of the approach is illustrated by means of examples in
 112 the field of ultrasound imaging, in which we also show quantitative comparisons with
 113 state-of-the art-methods for the joint deconvolution-segmentation task.

114
 115 The contributions of this work are (i) the proposition of an original variational
 116 model for the joint image recovery and features extraction problem; (ii) the design of
 117 a new block coordinate descent algorithm to address the resulting minimisation prob-
 118 lem; (iii) the convergence analysis of this scheme based on [6]; (iv) the illustration of
 119 the performance of the proposed method through two numerical examples in the field
 120 of image processing.

121
 122 The paper is organised as follows. In Section 2 we introduce the degradation
 123 model and report our derivation of the objective function for image recovery and
 124 feature extraction, starting from statistical assumptions on the data. In Section 3,
 125 we describe the proposed P-SASL-PAM method to address a general non-smooth
 126 non-convex optimization problem; secondly we show that the proposed method con-
 127 verges globally, in the sense that the whole generated sequence converges to a (local)
 128 minimum. The application of the P-SASL-PAM method to the joint reconstruction-
 129 segmentation problem is described in Section 4. Some illustrative numerical results
 130 are shown in Section 5. Conclusions are drawn in Section 6

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

137 **2.1. Observation Model.** Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be respectively the vec-
 138 torized sought-for solution and the observed data, which are assumed to be related
 139 according to the following model

140 (2.1)
$$y = Kx + \omega,$$

141 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 distri-
 142 bution with zero mean and covariance matrix $\sigma^2 \mathbb{I}_m$ with $\sigma > 0$ and \mathbb{I}_m states for the
 143 $m \times m$ identity matrix. We further assume that x can be characterised by a finite
 144 set of k features that are defined in a suitable space, where the data are described by
 145 a simple model relying on a small number of parameters. The Generalised Gaussian
 146 Distribution (\mathcal{GGD})

$$147 \quad (2.2) \quad (\forall t \in \mathbb{R}) \quad \mathfrak{p}(t; p, \alpha) = \frac{1}{2\alpha^{1/p} \Gamma\left(1 + \frac{1}{p}\right)} \exp\left(-\frac{|t|^p}{\alpha}\right), \quad (p, \alpha) \in (0, +\infty)^2$$

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

154 Assuming that K and σ are known, the task we address in this work is to jointly
 155 retrieve x (reconstruction) and obtain a good representation of its features through
 156 an estimation of the underlying model parameters (p_j, α_j) for $j \in \{1, \dots, k\}$ (fea-
 157 ture extraction). Starting from a similar statistical model as the one considered in
 158 [30, 74], we infer a continuous variational framework which does not rely on the *a pri-*
 159 *ori* knowledge of the exact number of features k . We derive this model by performing
 160 a *Maximum a Posteriori* estimation, which allows us to formulate the Joint Image
 161 Reconstruction and Feature Extraction task as a nonsmooth nonconvex optimisation
 162 problem involving a coupling term and a block-coordinate separable one.

163

164 **2.2. Bayesian Model.** From (2.1), we derive the following likelihood

$$165 \quad (2.3) \quad \mathfrak{p}(y|x, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|y - Kx\|^2}{2\sigma^2}\right).$$

166 Assuming then that the components of x are independent conditionally to the knowl-
 167 edge of their feature class, we define x as a mixture of \mathcal{GGDs}

$$168 \quad (2.4) \quad \mathfrak{p}(x|p, \alpha) = \prod_{j=1}^k \frac{1}{\left(2\alpha_j^{1/p_j} \Gamma\left(1 + \frac{1}{p_j}\right)\right)^{N_j}} \exp\left(-\frac{\|\bar{x}_j\|^{p_j}}{\alpha_j}\right).$$

169 Hereabove, for every $u \in \mathbb{R}^n$ and a feature labels set $j \in \{1, \dots, k\}$, we define $\bar{u}_j \in \mathbb{R}^{N_j}$
 170 as the vector containing only the N_j components of u that belong to the j -th feature.
 171 The shape parameters are assumed to be uniformly distributed on a certain interval
 172 $[a, b] \subset \mathbb{R}^{+*}$, while uninformative Jeffreys priors are assigned to the scale parameters:

$$173 \quad (2.5) \quad \mathfrak{p}(p) = \prod_{j=1}^k \frac{1}{b-a} \mathbf{1}_{[a,b]}(p_j),$$

174

$$175 \quad (2.6) \quad \mathfrak{p}(\alpha) = \prod_{j=1}^k \frac{1}{\alpha_j} \mathbf{1}_{\mathbb{R}^+}(\alpha_j).$$

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

178

179 **2.3. Variational Model.** In order to avoid to define *a priori* the number of
180 features, we regularise the problem by considering the 2D Total Variation (TV) of
181 the \mathcal{GGD} parameters $(p, \alpha) \in (0, +\infty)^n \times (0, +\infty)^n$. The idea of using Total Variation
182 to define a segmentation procedure is studied in [15, 16, 17, 19, 20, 55] by virtue of
183 the co-area formula: the authors propose to replace the boundary information term
184 of the Mumford-Shah (MS) functional [52] with the TV convex integral term. This
185 choice yields a nontight convexification of the MS model that does not require to
186 set the number of segments in advance. The overall segmentation procedure is then
187 built upon two steps: the first one consists of obtaining a smooth version of the given
188 image that is adapted to segmentation by minimising the proposed functional with
189 convex methods; the second step consists of partitioning the obtained solution into the
190 desired number of segments, by *e.g.* defining the thresholds with Otsu’s method [54]
191 or the k -means algorithm. The strength of our approach is that the second step (*i.e.*
192 the actual segmentation step) is independent from the first one, hence it is possible
193 to set the number of segments (*i.e.*, labels) without solving the optimisation problem
194 again.

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

$$199 \quad (2.7) \quad \beta_i = \frac{1}{p_i} \ln \alpha_i,$$

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

$$204 \quad (2.8) \quad \mathbf{p}(x|p, \beta) = \prod_{i=1}^n \frac{1}{2 \exp(\beta_i) \Gamma\left(1 + \frac{1}{p_i}\right)} \exp(-|x_i|^{p_i} \exp(-p_i \beta_i))$$

205

$$206 \quad (2.9) \quad \mathbf{p}(p) = \exp(-\lambda \text{TV}(p)) \prod_{i=1}^n \frac{1}{b-a} \mathbf{1}_{[a,b]}(p_i)$$

$$207 \quad (2.10) \quad \mathbf{p}(\beta) = \exp(-\zeta \text{TV}(\beta)) \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\left(-\frac{\beta_i^2}{2\sigma_\beta^2}\right).$$

208 The joint posterior distribution is determined as follows:

$$209 \quad \mathbf{p}(x, p, \beta|y) \propto \mathbf{p}(y|x, p, \beta) \mathbf{p}(x, p, \beta)$$

$$210 \quad (2.11) \quad \propto \mathbf{p}(y|x, p, \beta) \mathbf{p}(x|p, \beta) \mathbf{p}(p) \mathbf{p}(\beta).$$

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

215 and Features Extraction Problem:

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

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

$$222 \quad (2.13) \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2\sigma^2} \|y - Kx\|^2 + \sum_{i=1}^n |x_i|^{p_i} e^{-p_i \beta_i},$$

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

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

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

243 **3.1. Subdifferential Calculus.** Let us now recall some definitions and elements
244 of subdifferential calculus that will be useful in the upcoming sections. For a proper
245 and lower semicontinuous function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$, the domain of f is defined as

$$246 \quad \text{dom } f = \{u \in \mathbb{R}^n \mid f(u) < +\infty\}.$$

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

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

$$251 \quad \forall u \in \mathbb{R}^n \quad f(u) \geq f(u^+) + \langle v, u - u^+ \rangle.$$

252 *If $u^+ \notin \text{dom } f$, then $\partial f(u^+) = \emptyset$.*

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

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

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

259
$$(\forall u \in \mathbb{R}^n) \quad f(u) \geq f(u^+) + \langle v, u - u^+ \rangle + o(\|u - u^+\|);$$

260 *if $u^+ \notin \text{dom } f$, then $\hat{\partial}f(u^+) = \emptyset$;*

- 261 • *the limiting-subdifferential of f at u^+ , denoted by $\partial f(u^+)$, is defined as*

262
$$\partial f(u^+) = \left\{ u \in \mathbb{R}^n \mid \exists u^k \rightarrow u^+, f(u^k) \rightarrow f(u^+), v^k \rightarrow v, v^k \in \hat{\partial}f(u^k) \right\}.$$

263 If f is lower semicontinuous and convex, then the three previous notions of sub-
 264 differentiability are equivalent, i.e. $\hat{\partial}f(u^+) = \partial f(u^+)$. If f is differentiable, then
 265 $\partial f(u^+) = \{\nabla f(u^+)\}$. Now it is possible to formalise the notion of critical point
 266 for a general function:

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

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

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

273 (3.1)
$$(\forall u \in \mathbb{R}^n) \quad \|u\|_A = (u^\top A u)^{1/2}.$$

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

277 (3.2)
$$\text{prox}_f^A(u^+) = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left(\frac{1}{2} \|u - u^+\|_A^2 + f(u) \right).$$

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

281 **3.2. The KŁ-Property.** Most of the works related to BCD-based algorithms
 282 rely on the framework developed by Attouch, Bolte, and Svaiter in their seminal paper
 283 [6] in order to prove the convergence of block alternating strategies for nonsmooth and
 284 nonconvex problems. A fundamental assumption in [6] is that the objective function
 285 satisfies the *Kurdyka-Łojasiewicz* (KŁ) property [45, 48, 49]. We recall the definition
 286 of this property as it was given in [11]. Let $\eta \in (0, +\infty]$ and denote by Φ_η the class of
 287 concave continuous functions $\varphi : [0, +\infty) \rightarrow \mathbb{R}_+$ satisfying the following conditions:

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

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

$$293 \quad \text{dist}(u^+, S) = \inf_{u \in S} \|u^+ - u\|$$

294 with $\text{dist}(u^+, \emptyset) = +\infty$.

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

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

$$300 \quad (3.3) \quad f(u^+) < f(u) < f(u^+) + \eta \quad \Rightarrow \quad \varphi'(f(u) - f(u^+)) \text{dist}(0, \partial f(u)) \geq 1.$$

301 (ii) *Function f is said to be a KL function if it satisfies the KL property at each*
 302 *point of $\text{dom } \partial f$.*

303 **3.3. Proposed Algorithm.** Consider the general problem

$$304 \quad (3.4) \quad \underset{(x,p,\beta) \in (\mathbb{R}^n)^3}{\text{minimize}} \quad \left(\theta(x, p, \beta) = q(x, p, \beta) + f(x) + g(p) + h(\beta) \right)$$

305 associated to the following set of assumptions:

306 **ASSUMPTION 1.**

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

$$312 \quad (3.5) \quad \|\nabla q(x, p, \beta) - \nabla q(x^+, p^+, \beta^+)\| \\
 313 \quad \leq L_{q,S} (\|x - x^+\|^2 + \|p - p^+\|^2 + \|\beta - \beta^+\|^2)^{1/2}.$$

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

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

319 (iv) *θ is a KL function.*

320 We propose a block alternating algorithm to solve problem (3.4) which sequentially
 321 updates one of the three coordinate blocks (x, p, β) involved in function θ , through
 322 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 $\ell = 0, 1, \dots$

$$323 \quad (3.6) \quad x^{\ell+1} \in \text{prox}_{\gamma_1 q(\cdot, p^\ell, \beta^\ell)}^A(x^\ell - \gamma_1 A^{-1} \nabla f(x^\ell))$$

$$324 \quad (3.7) \quad p^{\ell+1} \in \text{prox}_{\gamma_2 \theta(x^{\ell+1}, \cdot, \beta^\ell)}(p^\ell)$$

$$325 \quad (3.8) \quad \beta^{\ell+1} \in \text{prox}_{\gamma_3 \theta(x^{\ell+1}, p^{\ell+1}, \cdot)}(\beta^\ell)$$

323 Our P-SASL-PAM approach is inspired from the rich literature on proximal ver-
 324 sions of BCD. We refer in particular to SLPAM [35] and ASAP [53] since our algorithm
 325 mixes both standard and linearised proximal regularisation on the coordinate blocks
 326 as in the former work, while inverting the splitting in order to gain more efficient
 327 proximal computations as in the latter. More precisely, we took advantage of the dif-
 328 ferentiability assumption on f to perform a linearised step for the update of variable
 329 x , while p and β are updated according to a standard proximal step. In addition,
 330 in order to accelerate the convergence, we introduced a preconditioned version of the
 331 linearised step which relies on the variable metric forward-backward strategy from
 332 [24]. As in [24], the preconditioning matrix $A \in \mathcal{S}_n$ is set so as to fulfill the following
 333 *majorisation condition*:

334 ASSUMPTION 2.

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

$$336 \quad (3.9) \quad (\forall x \in \mathbb{R}^n) \quad \phi(x, x^+) = f(x^+) + (x - x^+)^\top \nabla f(x^+) + \frac{1}{2} \|x^+ - x\|_A^2$$

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

$$338 \quad (3.10) \quad (\forall x \in \mathbb{R}^n) \quad f(x) \leq \phi(x, x^+).$$

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

$$340 \quad (3.11) \quad \underline{\nu} \mathbb{I}_n \preceq A \preceq \bar{\nu} \mathbb{I}_n.$$

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

$$343 \quad (\forall (x, x^+) \in \mathbb{R}^n \times \mathbb{R}^n) \quad f(x) \leq f(x^+) + (x - x^+)^\top \nabla f(x^+) + \frac{L_f}{2} \|x^+ - x\|^2.$$

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

347 **3.4. Convergence analysis.** In this subsection, we provide some technical re-
 348 sults regarding the sequences $(z^\ell)_{\ell \in \mathbb{N}} = ((x^\ell, p^\ell, \beta^\ell))_{\ell \in \mathbb{N}}$ and $(\theta(z^\ell))_{\ell \in \mathbb{N}}$ generated
 349 by Algorithm (3.1) that are instrumental to prove the convergence of the proposed
 350 method. Our proof relies on the general strategy designed in [6] which is based on
 351 three main ingredients: firstly a *sufficient decrease property*, secondly an *inexact opti-*
 352 *mality condition*, and finally the *Kurdyka-Łojasiewicz property*. On the one hand, this
 353 last property does not depend on the chosen algorithm, but only on the function at
 354 hand. In our framework, it is ensured by Assumption 1(iv). On the other hand, the
 355 first two properties, only related to the design of the algorithm itself, are expressed
 356 by Lemmas 3.7 and 3.9.

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

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

$$361 \quad (3.12) \quad \theta(z^{\ell+1}) \leq \theta(z^\ell) - \frac{\mu}{2} \|z^{\ell+1} - z^\ell\|^2.$$

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

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

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

$$366 \quad (3.13) \quad \text{prox}_{\gamma_1 q(\cdot, p^\ell, \beta^\ell)}^A(x^\ell - \gamma_1 A^{-1} \nabla f(x^\ell)) \\ 367 \quad = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left\{ q(u, p^\ell, \beta^\ell) + \frac{1}{2\gamma_1} \|u - x^\ell\|_A^2 + \langle \nabla f(x^\ell), u - x^\ell \rangle \right\}.$$

370 Hence,

$$371 \quad (3.14) \quad q(x^{\ell+1}, p^\ell, \beta^\ell) + \frac{1}{2\gamma_1} \|x^{\ell+1} - x^\ell\|_A^2 + \langle \nabla f(x^\ell), x^{\ell+1} - x^\ell \rangle \leq q(x^\ell, p^\ell, \beta^\ell). \\ 372$$

373 Moreover, the majorisation property (3.10) leads to

$$374 \quad (3.15) \quad f(x^{\ell+1}) \leq f(x^\ell) + \langle x^{\ell+1} - x^\ell, \nabla f(x^\ell) \rangle + \frac{1}{2} \|x^{\ell+1} - x^\ell\|_A^2.$$

375 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
376 exploit (3.15) together with (3.11), to obtain

$$377 \quad (3.16) \quad q(x^{\ell+1}, p^\ell, \beta^\ell) + f(x^{\ell+1}) + \frac{\nu}{2} \left(\frac{1}{\gamma_1} - 1 \right) \|x^{\ell+1} - x^\ell\|^2 \leq q(x^\ell, p^\ell, \beta^\ell) + f(x^\ell), \\ 378$$

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

380 Then, the variational definition of the proximal steps on functions p and β implies
381 that
382

$$383 \quad (3.17) \quad \frac{1}{2\gamma_2} \|p^{\ell+1} - p^\ell\|^2 + g(p^{\ell+1}) + q(x^{\ell+1}, p^{\ell+1}, \beta^\ell) \leq g(p^\ell) + q(x^{\ell+1}, p^\ell, \beta^\ell),$$

384

$$385 \quad (3.18) \quad \frac{1}{2\gamma_3} \|\beta^{\ell+1} - \beta^\ell\|^2 + h(\beta^{\ell+1}) + q(x^{\ell+1}, p^{\ell+1}, \beta^{\ell+1}) \leq h(\beta^\ell) + q(x^{\ell+1}, p^{\ell+1}, \beta^\ell).$$

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

$$387 \quad \frac{\nu}{2} \left(\frac{1}{\gamma_1} - 1 \right) \|x^{\ell+1} - x^\ell\|^2 + \frac{1}{2\gamma_2} \|p^{\ell+1} - p^\ell\|^2 + \frac{1}{2\gamma_3} \|\beta^{\ell+1} - \beta^\ell\|^2 \\ 388 \quad (3.19) \quad \leq \theta(x^\ell, p^\ell, \beta^\ell) - \theta(x^{\ell+1}, p^{\ell+1}, \beta^{\ell+1}).$$

390 Thus, by setting $(z^\ell)_{\ell \in \mathbb{N}} = ((x^\ell, p^\ell, \beta^\ell))_{\ell \in \mathbb{N}}$ and defining the positive constant
391 $\mu = \min\{\frac{\nu}{2}(\frac{1}{\gamma_1} - 1), \frac{1}{\gamma_2}, \frac{1}{\gamma_3}\}$, we get (3.12).

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

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

$$395 \quad (3.20) \quad \sum_{\ell=0}^N \|z^\ell - z^{\ell+1}\|^2 \leq \frac{1}{\mu} \sum_{\ell=0}^N (\theta(z^\ell) - \theta(z^{\ell+1}))$$

$$396 \quad (3.21) \quad = \frac{1}{\mu} (\theta(z^0) - \theta(z^{N+1}))$$

$$397 \quad (3.22) \quad \leq \frac{1}{\mu} (\theta(z^0) - \underline{\theta}).$$

398

399 Taking the limit as $N \rightarrow +\infty$ yields the desired summability property. \square

400 Before presenting the *inexact optimality* property for any sequence generated by the
401 proposed method, we recall an important result regarding function θ appearing in
402 (3.4), under Assumption 1:

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

$$405 \quad \partial\theta(x, p, \beta)$$

$$406 \quad = \partial_x\theta(x, p, \beta) \times \partial_p\theta(x, p, \beta) \times \partial_\beta\theta(x, p, \beta)$$

$$407 \quad = \{\nabla_x q(x, p, \beta) + \nabla f(x)\} \times (\nabla_p q(x, p, \beta) + \partial g(p)) \times (\nabla_\beta q(x, p, \beta) + \partial h(\beta)).$$

408

409

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

$$412 \quad (3.23) \quad \|b^\ell\| \leq \rho \|z^{\ell-1} - z^\ell\|,$$

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

414 *Proof.* The assumed boundedness obviously implies that there exists a bounded
415 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
416 $\{(x^\ell, p^\ell, \beta^{\ell-1})\}_{\ell \in \mathbb{N}}$ are included in S . In addition, recall that, according to Assump-
417 tion 1, the coupling term q has a Lipschitz continuous gradient on S . In the following,
418 we will exploit the fact that, at every iteration ℓ , the update for each block of coord-
419 inate needs to satisfy Fermat's rule for the corresponding subproblem.

420

421 • Fermat's rule for (3.6) reads

$$422 \quad (3.24) \quad \gamma_1^{-1} A(x^{\ell-1} - x^\ell) = \nabla f(x^{\ell-1}) + \nabla_x q(x^\ell, p^{\ell-1}, \beta^{\ell-1}).$$

423 Notice that

$$424 \quad (3.25) \quad \partial_x\theta(x^\ell, p^\ell, \beta^\ell) = \{\nabla_x\theta(x^\ell, p^\ell, \beta^\ell)\} = \{\nabla f(x^\ell) + \nabla_x q(x^\ell, p^\ell, \beta^\ell)\}.$$

425 So, by defining

$$426 \quad (3.26) \quad b_x^\ell = \gamma_1^{-1} A(x^{\ell-1} - x^\ell) + \nabla f(x^\ell) + \nabla_x q(x^\ell, p^\ell, \beta^\ell)$$

$$427 \quad - \nabla f(x^{\ell-1}) - \nabla_x q(x^\ell, p^{\ell-1}, \beta^{\ell-1}),$$

428

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

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

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

- Fermat's rule for (3.7) reads

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

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

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

by defining

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

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

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

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

- Fermat's rule for (3.8) reads

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

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

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

and defining

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

we have

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

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). \square

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

463 **PROPOSITION 3.10** (Properties of the cluster points set). *Suppose that Assump-*
 464 *tions 1 and 2 hold. Let $(z^\ell)_{\ell \in \mathbb{N}}$ be a sequence generated by Algorithm 3.1. Denote by*
 465 *$\omega(z^0)$ 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*

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

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

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

$$468 \quad \inf_{\ell \in \mathbb{N}} \theta(z^\ell) = \lim_{\ell \rightarrow +\infty} \theta(z^\ell).$$

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

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

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

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

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

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

$$483 \quad (3.36) \quad (\forall X \in \mathbb{R}^N) \quad \theta(X) = q(X) + \sum_{k=1}^S g_k(X_k).$$

484 Hereabove, $X = (X_1, \dots, X_S) \in \mathbb{R}^N$, with each $X_k \in \mathbb{R}^{n_k}$, $k \in \{1, \dots, S\}$, so
 485 that $N = \sum_{k=1}^S n_k$. Function θ involves a a locally Lipschitz-differentiable cou-
 486 pling term $q: \mathbb{R}^N \rightarrow \mathbb{R}$ and S block-separable terms $g_k: \mathbb{R}^{n_k} \rightarrow]-\infty, +\infty[$ (with
 487 $k \in \{1, \dots, S\}$), some of which may be differentiable with a Lipschitz continuous gra-
 488 dient. Then, the generalized variant of P-SASL-PAM generates a sequence $(Z^\ell)_{\ell \in \mathbb{N}} =$
 489 $((X_1^\ell, \dots, X_S^\ell))_{\ell \in \mathbb{N}}$ where the blocks of coordinates are updated via the following
 490 scheme, at every iteration $\ell \in \mathbb{N}$:

$$491 \quad \begin{cases} \text{For } k = 1, \dots, S \\ \left\{ \begin{array}{l} X_k^{\ell+1} \in \text{prox}_{\gamma_k q(X_1^{\ell+1}, \dots, X_{k-1}^{\ell+1}, \cdot, X_{k+1}^\ell, \dots, X_S^\ell)}^{A_k}(X_k^\ell - \gamma_k A_k^{-1} \nabla g_k(X_k^\ell)) \\ \quad \text{with } \gamma_k \in (0, 1), \text{ if } g_k \text{ is differentiable,} \\ X_k^{\ell+1} \in \text{prox}_{\gamma_k \theta(X_1^{\ell+1}, \dots, X_{k-1}^{\ell+1}, \cdot, X_{k+1}^\ell, \dots, X_S^\ell)}(X_k^\ell) \\ \quad \text{with } \gamma_k > 0, \text{ otherwise.} \end{array} \right. \end{cases}$$

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

494 *Decrease* and the *Inexact Optimality* properties expressed in Lemma 3.7 and Lemma
 495 3.9 can be extended to a bounded sequence $(Z^\ell)_{\ell \in \mathbb{N}}$ generated by the above extended
 496 variant of P-SASL-PAM. In addition, if function θ is a KL function, then the con-
 497 vergence results expressed in Proposition 3.11 can be extended, ensuring that the
 498 above mentioned sequence has a finite length, *i.e.* $\sum_{\ell=1}^{\infty} \|Z^{\ell+1} - Z^\ell\| < +\infty$, and it
 499 converges to a critical point Z^* of θ .

500 4. Application of P-SASL-PAM to the Joint Reconstruction and Fea- 501 ture Extraction Problem.

502 **4.1. Smoothing of the coupling term.** The application of Algorithm 3.1 to
 503 Problem (2.12) requires the involved functions to fulfill the requirements listed in
 504 Assumption 1. This section is devoted to this analysis, by first defining the following
 505 functions, for every $x = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, $p = (p_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, and $\beta = (\beta_i)_{1 \leq i \leq n} \in$
 506 \mathbb{R}^n ,

$$507 \quad (4.1) \quad \tilde{q}(x, p, \beta) = \sum_{i=1}^n |x_i|^{p_i} e^{-\beta_i p_i},$$

$$508 \quad (4.2) \quad f(x) = \frac{1}{2\sigma^2} \|y - Kx\|_2^2,$$

$$509 \quad (4.3) \quad g(p) = \sum_{i=1}^n \left(\ln \Gamma\left(1 + \frac{1}{p_i}\right) + \iota_{[a,b]}(p_i) \right) + \lambda \text{TV}(p),$$

$$510 \quad (4.4) \quad h(\beta) = \sum_{i=1}^n \left(\beta_i + \frac{\beta_i^2}{2\sigma_\beta^2} \right) + \zeta \text{TV}(\beta).$$

512 The first item in Assumption 1 regarding the regularity of the coupling term is not
 513 satisfied by (4.1). To circumvent this difficulty, we introduce the *pseudo-Huber loss*
 514 *function* [21] depending on a pair of parameters $\delta = (\delta_1, \delta_2) \in (0, +\infty)^2$ such that
 515 $\delta_2 < \delta_1$:

$$516 \quad (4.5) \quad (\forall t \in \mathbb{R}) \quad C_\delta(t) = H_{\delta_1}(t) - \delta_2,$$

517 where H_{δ_1} is the *hyperbolic function* defined, for every $t \in \mathbb{R}$, by $H_{\delta_1}(t) = \sqrt{t^2 + \delta_1^2}$.
 518 Function (4.5) is used as a smooth approximation of the absolute value involved in
 519 (4.1). We then replace (4.1) with

$$520 \quad (4.6) \quad q(x, p, \beta) = \sum_{i=1}^n (C_\delta(x_i))^{p_i} e^{-\beta_i p_i}.$$

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

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

527

528 Now, we must show that Θ is a KL function. To do so, let us consider the notion
 529 of *o-minimal structure* [66], which is a particular family $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ where each \mathcal{O}_n
 530 is a collection of subsets of \mathbb{R}^n , satisfying a series of axioms (we refer to [5, Definition
 531 13], for a complete illustration). We present hereafter the definition of *definable set*
 532 and *definable function* in an o-minimal structure:

533 DEFINITION 4.1 (Definable sets and definable functions). *Given an o-minimal*
 534 *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*
 535 *extended valued function $f : \mathbb{R} \rightarrow (-\infty, +\infty]$ is said to be definable in \mathcal{O} if its graph*
 536 *is a definable subset of $\mathbb{R}^n \times \mathbb{R}$.*

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

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

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

556

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

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

560 *Proof.* Let us show that there exists $\mu > 0$ such that function $t \mapsto \ln \Gamma(1 +$
 561 $\frac{1}{t}) + \mu t^2/2$ is convex on $(0, +\infty)$. The second-order derivative of this function on the
 562 positive real axis reads

$$563 \quad \frac{d^2}{dt^2} \left(\ln \Gamma \left(1 + \frac{1}{t} \right) + \frac{\mu}{2} t^2 \right)$$

$$564 \quad (4.7) \quad = \frac{1}{t^3} \left(2 \text{Digamma} \left(1 + \frac{1}{t} \right) + \frac{1}{t} \text{Digamma}' \left(1 + \frac{1}{t} \right) + \mu t^3 \right),$$

565

566 where the Digamma function is the logarithmic derivative of the Gamma function.
 567 In order to show the convexity of the considered function, we need to ensure that
 568 (4.7) is positive for every $t \in (0, +\infty)$. By virtue of Bohr-Möllerup's theorem
 569 [4, Theorem 2.1], among all functions extending the factorial functions to the posi-
 570 tive real numbers, only the Gamma function is log-convex. More precisely, its
 571 natural logarithm is (strictly) convex on the positive real axis. This implies that
 572 $t \mapsto \text{Digamma}'(t)$ is positive. It results that the only sign-changing term in (4.7)
 573 is function $t \mapsto 2 \text{Digamma} \left(1 + \frac{1}{t} \right)$ as $t \mapsto \text{Digamma}(t)$ vanishes in a point $t_0 > 1$
 574 ($t_0 \approx 1.46163$) which corresponds to the minimum point of the Gamma function – and
 575 therefore also of its natural logarithm [72]. As a consequence, the Digamma function
 576 is strictly positive for $t \in (t_0, +\infty)$, implying that $t \mapsto \text{Digamma} \left(1 + \frac{1}{t} \right)$ is strictly
 577 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

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

$$579 \quad \frac{d}{dt} \text{Digamma} \left(1 + \frac{1}{t} \right) = -\frac{1}{t^2} \text{Digamma}' \left(1 + \frac{1}{t} \right)$$

580 and by the following limit

$$581 \quad \lim_{t \rightarrow +\infty} \text{Digamma} \left(1 + \frac{1}{t} \right) = \text{Digamma}(1) = -\mathcal{E}$$

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

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

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

583 **4.2. Proximal computations.** Let us now focus on the proximal computations
584 involved in Algorithm 3.1. Given the elaborate structure of the involved functions,
585 no trivial closed-form expression is available to compute the required proximity op-
586 erators. Luckily, efficient minimisation strategies can be designed to tackle the three
587 inner optimisation problems. To ease the description, we summarize in Algorithm 4.1
588 the application of Algorithm 3.1 to the resolution of (2.12).
589

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 $\ell = 0, 1, \dots$

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

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

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

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

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

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

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

$$\begin{aligned} 596 \quad (\forall u > \delta_2) \quad (u - \delta_2)^p &\leq (w - \delta_2)^p + p(w - \delta_2)^{p-1}(u - w), \\ 597 \quad (4.11) \quad &= (1 - p)(w - \delta_2)^p + p(w - \delta_2)^{p-1}(u - \delta_2). \end{aligned}$$

599 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
600 the following majorisation:

$$601 \quad (4.12) \quad (C_\delta(t))^p \leq (1 - p)(C_\delta(t'))^p + p(C_\delta(t'))^{p-1}C_\delta(t).$$

603 Let us now define $\mathcal{I}^\ell = \{i \in \{1, \dots, n\} \mid p_i^\ell \geq 1\}$ and $\mathcal{J}^\ell = \{1, \dots, n\} \setminus \mathcal{I}^\ell$. Given
604 $v = (v_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, we deduce from (4.12) that

$$\begin{aligned} 605 \quad (\forall x = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad q(x, p^\ell, \beta^\ell) &= \sum_{i \in \mathcal{I}^\ell} (C_\delta(x_i))^{p_i^\ell} e^{-\beta_i^\ell p_i^\ell} + \sum_{i \in \mathcal{J}^\ell} (C_\delta(x_i))^{p_i^\ell} e^{-\beta_i^\ell p_i^\ell} \\ 606 \quad (4.13) \quad &\leq \bar{q}(x, v, p^\ell, \beta^\ell), \end{aligned}$$

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

$$609 \quad (4.14) \quad \bar{q}(x, v, p^\ell, \beta^\ell) = \sum_{i=1}^n \bar{q}_i(x_i, v_i, p_i^\ell, \beta_i^\ell),$$

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

$$\begin{aligned} 611 \quad (4.15) \quad \bar{q}_i(x_i, v_i, p_i^\ell, \beta_i^\ell) &= \begin{cases} e^{-\beta_i^\ell p_i^\ell} (C_\delta(x_i))^{p_i^\ell}, & \text{if } p_i^\ell \geq 1 \\ e^{-\beta_i^\ell p_i^\ell} \left((C_\delta(v_i))^{p_i^\ell} (1 - p_i^\ell) + p_i^\ell (C_\delta(v_i))^{p_i^\ell - 1} C_\delta(x_i) \right) & \text{otherwise.} \end{cases} \\ 612 \quad & \\ 613 \quad & \end{aligned}$$

614 In a nutshell, each term of index $i \in \{1, \dots, n\}$ in (4.14) coincides either with the i -th
615 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
616 to v_i when $i \in \mathcal{J}^\ell$. We thus propose to adopt a majorisation-minimisation procedure
617 by building a sequence of convex surrogate problems for the nonconvex minimisation
618 problem involved in the computation of $\text{prox}_{\gamma_1 q(\cdot, p^\ell, \beta^\ell)}^A$. At the κ -th iteration of this
619 procedure, following the MM principle, the next iterate $x^{\kappa+1}$ is determined by setting
620 $v = x^\kappa$. We summarise the strategy in Algorithm 4.2.

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

Initialize $x^0 \in \mathbb{R}^n$

For $\kappa = 0, 1, \dots$ **until** convergence

$$(4.16) \quad x^{\kappa+1} = \text{prox}_{\gamma_1 \bar{q}(\cdot, x^\kappa, p^\ell, \beta^\ell)}^A(x^+) \quad (\text{with Alg. 4.3}).$$

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

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

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

Initialize dual variable $w^0 \in \mathbb{R}^n$

Set $\eta \in (0, 2\|A\|^{-1})$

For $\kappa' = 0, 1, \dots$ **until** convergence

$$(4.17) \quad u^{\kappa'} = x^+ - Aw^{\kappa'},$$

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

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

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

628

$$629 \quad (4.19) \quad (\forall u^+ = (u_i^+)_{1 \leq i \leq n} \in \mathbb{R}^n)$$

$$630 \quad \text{prox}_{\eta^{-1}\gamma_1 \bar{q}(\cdot, v, p^\ell, \beta^\ell)}(u^+) = \left(\text{prox}_{\eta^{-1}\gamma_1 \bar{q}_i(\cdot, v_i, p_i^\ell, \beta_i^\ell)}(u_i^+) \right)_{1 \leq i \leq n}.$$

631

632 More precisely,

633

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

634

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

635

636

637 The proximity operator of the so-scaled version of function H_{δ_1} can be deter-
638 mined by solving a quartic polynomial equation.¹

639

- For every $i \in \{1, \dots, n\}$ such that $p_i^\ell > 1$,

640

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

641

642

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

643

644

645

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.

646

647

648

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

649

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

¹<http://proximity-operator.net/scalarfunctions.html>

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

651

652 (4.23) $(\forall i \in \{1, \dots, n\})(\forall p_i \in \mathbb{R})$

$$653 \quad \psi_i^\ell(p_i) = \begin{cases} (C_\delta(x_i^{\ell+1}))^{p_i} e^{-\beta_i^\ell p_i} + \ln \Gamma(1 + \frac{1}{p_i}) + \frac{1}{2\gamma_2}(p_i - p_i^\ell)^2 & \text{if } p_i > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

654

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

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

655 Problem (4.22) is equivalent to minimizing the sum of the indicator function of
656 a hypercube, a separable component, and a nonseparable term involving the linear
657 operator D . According to Proposition 4.3, we can ensure the convexity of each term
658 $(\psi_i^\ell)_{1 \leq i \leq n}$ by setting $\gamma_2 < \frac{1}{\mu_0} \approx 8.805$. In order to solve (4.22), it is then possible to
659 implement a Primal-Dual (PD) algorithm [29, 44, 69] as outlined in Algorithm 4.4.

660

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

$$(4.24) \quad u^\kappa = p^\kappa - \tau(D^* v_1^\kappa + v_2^\kappa),$$

$$(4.25) \quad p^{\kappa+1} = \text{proj}_{[a,b]^n}(u^\kappa),$$

$$(4.26) \quad w_1^\kappa = v_1^\kappa + \sigma D(2p^{\kappa+1} - p^\kappa),$$

$$(4.27) \quad v_1^{\kappa+1} = w_1^{\kappa+1} - \sigma \text{prox}_{\frac{\lambda \ell_{1,2}}{\sigma}}\left(\frac{w_1^\kappa}{\sigma}\right).$$

$$(4.28) \quad w_2^\kappa = v_2^\kappa + \sigma(2p^{\kappa+1} - p^\kappa),$$

$$(4.29) \quad v_2^{\kappa+1} = w_2^{\kappa+1} - \sigma \text{prox}_{\frac{\psi^\ell}{\sigma}}\left(\frac{w_2^\kappa}{\sigma}\right).$$

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

661 The proximity operator of the involved $\ell_{1,2}$ norm has a closed-form expression.

662 For every $w_1 = ([w_1]_{i,1}, [w_1]_{i,2})_{1 \leq i \leq n} \in \mathbb{R}^{n \times 2}$ and $\lambda > 0$, we have

$$663 \quad \text{prox}_{\lambda \ell_{1,2}}(w_1) = \left(\text{prox}_{\lambda \|\cdot\|_2} \left(([w_1]_{i,1}, [w_1]_{i,2}) \right) \right)_{1 \leq i \leq n}$$

$$664 \quad = \left(([w_1]_{i,1}, [w_1]_{i,2}) - \frac{\lambda([w_1]_{i,1}, [w_1]_{i,2})}{\max\{\lambda, \|([w_1]_{i,1}, [w_1]_{i,2})\|_2\}} \right)_{1 \leq i \leq n}.$$

665

The proximal point at $w_2^\kappa/\sigma = ([w_2^\kappa]_i/\sigma)_{1 \leq i \leq n} \in \mathbb{R}^n$ 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

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

The update in (4.29) then reads

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

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

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

666

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

$$669 \quad (4.30) \quad \underset{\beta \in \mathbb{R}^n}{\text{minimize}} \quad \varphi^\ell(\beta) + \zeta \ell_{1,2}(D\beta)$$

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 $i \in \{1, \dots, n\}$,

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

670 The above problem shares structure similar to the one studied in the previous case
671 since the objective function is the sum of the smooth convex term φ^ℓ and the nons-
672 mooth convex one $\zeta \text{TV} = \zeta \ell_{1,2}(D\cdot)$, and it can be solved by the primal-dual procedure
673 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 $\kappa = 0, 1, \dots$ **until** convergence

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

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

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

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

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

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

$$675 \quad (4.35) \quad (\forall \beta = (\beta_i)_{1 \leq i \leq n} \in \mathbb{R}^n) \quad \text{prox}_{\tau\varphi^\ell}(\beta) = \left(\text{prox}_{\tau\varphi_i^\ell}(\beta_i) \right)_{1 \leq i \leq n}$$

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

$$677 \quad (4.36) \quad (\forall \beta_i \in \mathbb{R}) \quad \mathbf{h}_i(\beta_i) = \varphi_i^\ell(\beta_i) + \frac{1}{2\tau}(\beta_i - u_i^\kappa)^2.$$

679 The nonlinear equation defining the unique zero of the derivative of \mathbf{h}_i admits a closed-
680 form solution that involves the Lambert W -function [31]. Indeed, let us introduce the
681 following notation:

$$682 \quad (4.37) \quad a_{1,i} = p_i^{\ell+1} (C_\delta(x_i^{\ell+1}))^{p_i^{\ell+1}},$$

$$683 \quad (4.38) \quad a_2 = \left(\frac{1}{\sigma_\beta^2} + \frac{1}{\gamma_3} + \frac{1}{\tau} \right)^{-1},$$

$$684 \quad (4.39) \quad a_{3,i} = 1 - \frac{\beta_i^\ell}{\gamma_3} - \frac{u_i^\kappa}{\tau}.$$

686 Then

$$687 \quad \mathbf{h}'_i(\beta_i) = 0 \iff -a_{1,i} \exp(-p_i^{\ell+1}\beta_i) + \frac{\beta_i}{a_2} + a_{3,i} = 0$$

$$688 \quad \iff p_i^{\ell+1}(\beta_i + a_2 a_{3,i}) \exp(p_i^{\ell+1}(\beta_i + a_2 a_{3,i})) = p_i^{\ell+1} a_{1,i} a_2 \exp(p_i^{\ell+1} a_2 a_{3,i})$$

$$689 \quad (4.40) \quad \iff \beta_i = \frac{1}{p_i^{\ell+1}} W(p_i^{\ell+1} a_{1,i} a_2 \exp(p_i^{\ell+1} a_2 a_{3,i})) - a_2 a_{3,i},$$

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

$$693 \quad (4.41) \quad X \exp(X) = Y \iff X = W(Y).$$

694 In conclusion, the update in (4.32) reads as $\beta_i^{\kappa+1} = (\beta_i^{\kappa+1})_{1 \leq i \leq n}$ where each compo-
695 nent of this vector is calculated according to (4.40).

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

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

706 The model parameters that need to be tuned are the regularisation parameters
707 $(\lambda, \zeta) \in]0, +\infty[^2$ for the TV terms, the $\delta_1 > 0$ and $\delta_2 > 0$ values for the pseudo Huber
708 function, and the standard deviation $\sigma_\beta > 0$ for the reparameterised scale parameter.
709 Parameters (λ, ζ) are identified via an empirical search based on visual inspection,
710 considering the fact that the higher the effect of the TV regularisation term, the
711 flatter the estimated solution is. The third parameter, $\delta = (\delta_1, \delta_2)$, is tuned so that
712 δ_1 is chosen in the range $[0.1, 3]$ as the one that defines the best trade-off between a
713 high PSNR and a high overall accuracy (OA) value, while $\delta_2 = \delta_1 - 0.01$. Eventually,
714 for the last parameter σ_β , experimental results give credit to the fact that the choice
715 $\sigma_\beta = 1$ 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

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

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

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

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

$$728 \quad \text{PSNR} = 10 \log_{10} \left(n \max_{i \in \{1, \dots, n\}} (x_i, \hat{x}_i)^2 / \|x_i - \hat{x}_i\|^2 \right),$$

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

731 The stopping criterions for both the P-SASL-PAM outer loop and the inner loops
732 are set by defining a threshold level on the relative change between two consecutive
733 iterates of the involved variables and a maximum number of iterations. The outer
734 loop in Algorithm 3.1 stops whenever $\ell = 2000$ or when $\|z^{\ell+1} - z^\ell\| / \|z^\ell\| < 10^{-4}$.
735 The MM procedure to compute $x^{\ell+1}$ in Algorithm 4.2 is stopped after 300 iterations
736 or when $\|x^{\kappa+1} - x^\kappa\| / \|x^\kappa\| < 5 \times 10^{-3}$. The DFB procedure in Algorithm 4.3 to
737 compute $u^{\kappa+1}$ is stopped after 300 iterations or when $\|u^{\kappa+1} - u^\kappa\| / \|u^\kappa\| < 10^{-3}$. The
738 PD procedure in Algorithms 4.4 and 4.5 computing $p^{\ell+1}$ (resp. $\beta^{\ell+1}$) terminates af-
739 ter 200 iteration or when $\|p^{\kappa+1} - p^\kappa\| / \|p^\kappa\| < 10^{-3}$ (resp. $\|\beta^{\kappa+1} - \beta^\kappa\| / \|\beta^\kappa\| < 10^{-3}$).
740

741 In the first example, we illustrate the performance of the proposed method on an
742 image that is composed of three different texture regions. For the second example,
743 we work on the ultrasound images considered in [30, Section IV.C] and provide a
744 quantitative comparison with the methods that are mentioned in this work - namely,
745 a combination of Wiener deconvolution and Otsu’s segmentation [54], a combination
746 of Lasso deconvolution and SLaT segmentation [15], the adjusted Hamiltonian Monte
747 Carlo (HMC) method [60], the Proximal Unadjusted Langevin algorithm (P-ULA)
748 [56] and its preconditioned version (PP-ULA) [30] for joint deconvolution and seg-
749 mentation.

750 **5.1. Example 1.** In this first example, we propose an illustration of the ability
751 of the proposed framework to deconvolve and segment a synthetically created image.
752 The image, which we refer to as *Texture*, is a combination of three textures belonging
753 to the Original Brodatz’s Database² [1]. Each texture is located in a distinct region,

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

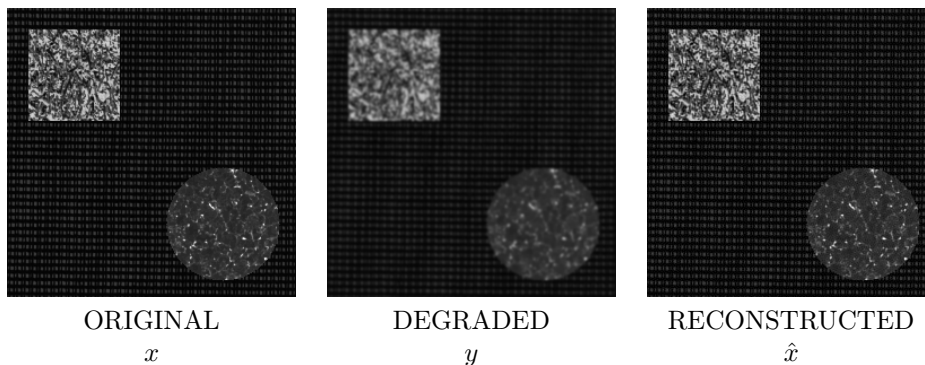


FIG. 1. Original (x), Degraded (y) and Reconstructed (\hat{x}) versions of Texture.

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

763 As a starting point for the algorithm we choose x^0 as the degraded image y ,
 764 $(p_i^0)_{1 \leq i \leq n}$ is drawn from an i.i.d. uniform random distribution over the range $[0.5, 1.5]$,
 765 and $(\beta_i^0)_{1 \leq i \leq n}$ is drawn from an i.i.d Gaussian distribution with zero mean and stan-
 766 dard deviation $\sigma_\beta = 1$. Figure 1(right) the reconstructed image \hat{x} , using P-SASL-
 767 PAM. Figure 2 shows the ability of P-SASL-PAM to accurately reconstruct a piecewise
 768 constant approximation of the shape parameter p . Namely, we propose a compari-
 769 son between the reference labelling of the textures in the image (Fig. 2(left)) and
 770 the estimated shape parameter \hat{p} (Fig. 2(middle)) along with its quantised version \bar{p}
 771 (Fig. 2(right)), which corresponds to our estimated labelling. Figure 3 shows the de-
 772 cay of the cost function θ along the first 500 iterations of P-SASL-PAM, assessing its
 773 fast and stable convergence. Eventually, we report in Table 1 the metrics of the quan-
 774 titative evaluation of the estimated solution, which confirms the good performance of
 775 our proposed method.

PSNR	SSIM	OA
33.13	0.95	99.7

TABLE 1
PSNR, SSIM and OA for Texture

776 **5.2. Example 2.** In this example, we illustrate the good performance of our
 777 approach on the joint deconvolution/segmentation of realistically simulated ultra-
 778 sound images with two regions (*Simu1*) and three regions (*Simu2*) extracted from
 779 [30]. We define K as the linear operator modelling the convolution with the point
 780 spread function of the probe, and set the noise variance to $\sigma^2 = 0.013$ for *Simu1* and
 781 $\sigma^2 = 33$ for *Simu2*. Following the procedure outlined in [30], we initialise x^0 using

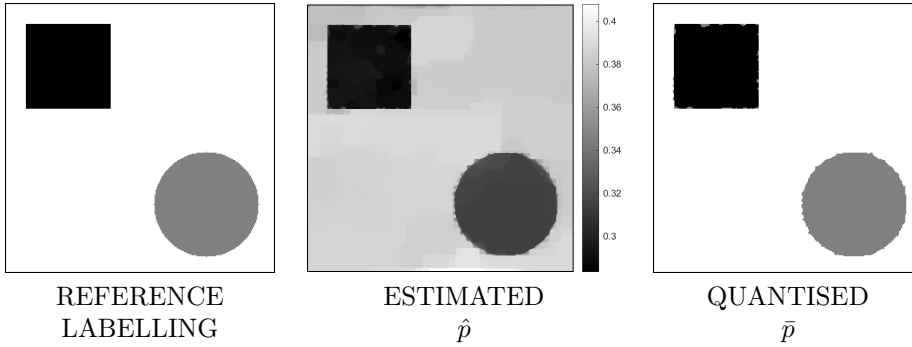


FIG. 2. Segmentation of the shape parameter for Texture: reference labelling, estimated \hat{p} and quantised \bar{p} .

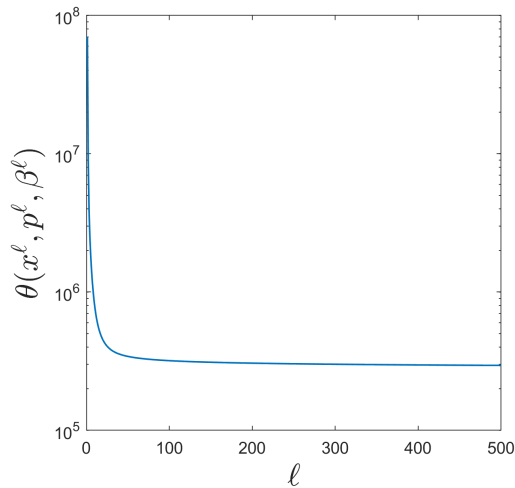


FIG. 3. Texture: Decay of the objective value along 500 iterations.

782 a pre-deconvolved image obtained with a Wiener filter applied to the observed data
 783 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
 784 $(\beta_i^0)_{1 \leq i \leq n}$ is drawn from an i.i.d. Gaussian distribution with zero mean and unit
 785 standard deviation.

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

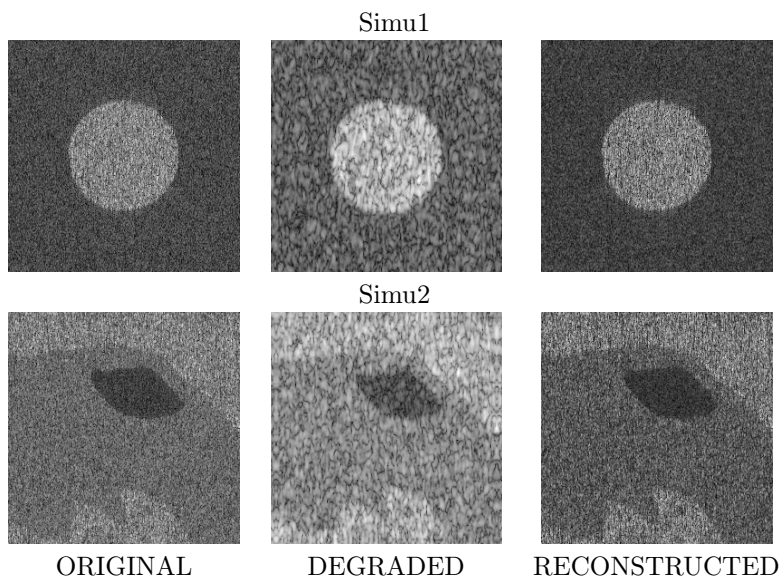


FIG. 4. *B-mode of Simu1 and Simu2. The B-mode image is the most common type of ultrasound image, displaying the acoustic impedance of a 2-dimensional cross section of the considered tissue. All images are presented in the same scale [0,1].*

799 Figure 7 shows the evolution of the cost function for both Simu1 and Simu2 along
800 1000 iterations, hereagain showing the great convergence behavior of our algorithm.

801 Eventually, Tables 2 and 3 propose a quantitative comparison of our results
802 against those of the methods considered in [30]. From these tables we can conclude
803 that the proposed variational method is able to compete with state-of-the-art Monte
804 Carlo Markov Chain techniques in terms of both segmentation and deconvolution
805 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

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

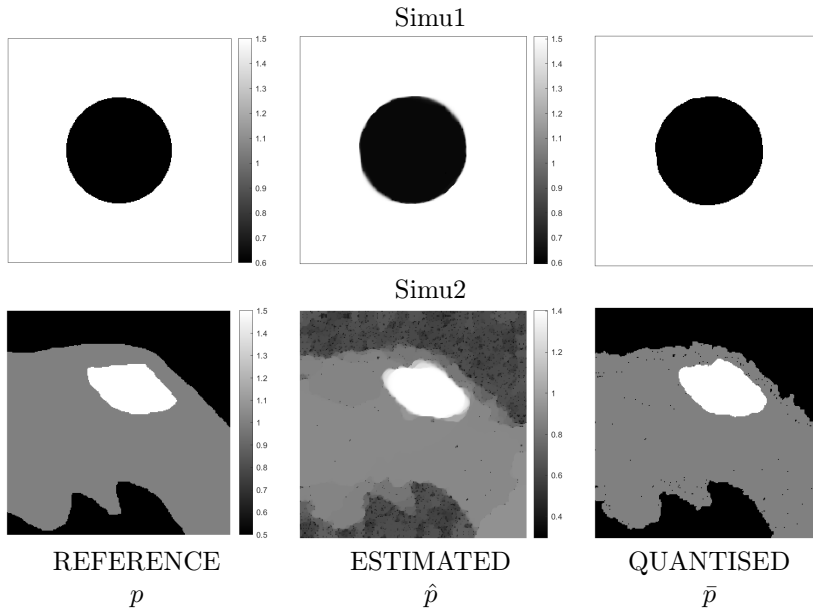


FIG. 5. Segmentation of the shape parameter for Simu1 and Simu2: reference p , estimated \hat{p} and quantised \bar{p} .

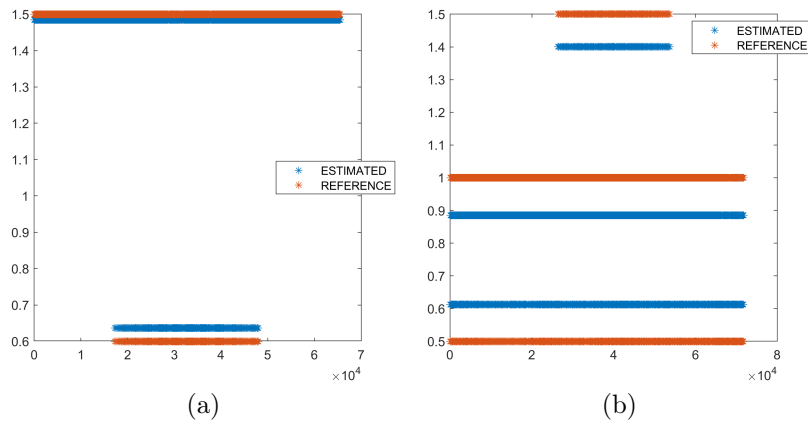


FIG. 6. Plot of the vectorised estimated shape parameter median values (blue) against the reference values (red) for Simu1 (a) and Simu2 (b).

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

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

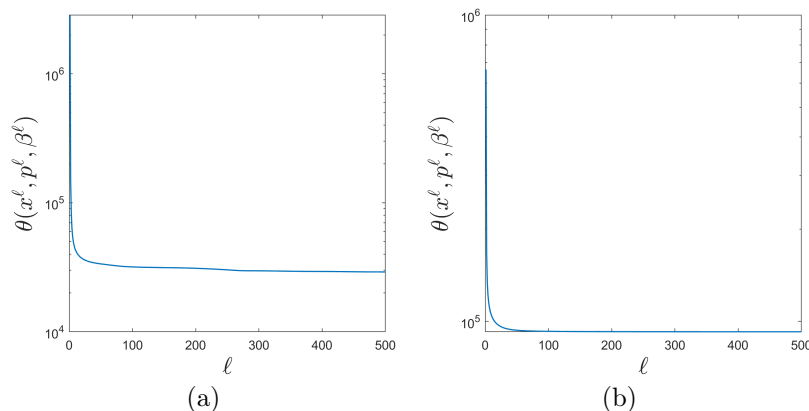


FIG. 7. Decay of the objective value along 500 iterations for Simu1 (a) and Simu2 (b).

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

824

REFERENCES

825 [1] S. ABDELMOUNAIME AND H. DONG-CHEN, *New Brodatz-based image databases for grayscale*
826 *color and multiband texture analysis*, International Scholarly Research Notices, 2013 (2013).
827 [2] M. AHOOKHOSH, L. T. K. HIEN, N. GILLIS, AND P. PATRINOS, *Multi-block Bregman proximal*
828 *alternating linearized minimization and its application to orthogonal nonnegative matrix*
829 *factorization*, Computational Optimization and Applications, 79 (2021), p. 681–715.
830 [3] G. E. ANDREWS, R. ASKEY, AND R. ROY, *Special Functions*, Encyclopedia of Mathematics
831 and its Applications, Cambridge University Press, 1999.
832 [4] E. ARTIN, *The Gamma Function*, Athena series, Holt, Rinehart and Winston, 1964.
833 [5] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimiza-*
834 *tion and projection methods for nonconvex problems: an approach based on the Kurdyka-*
835 *Lojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), p. 438–457.
836 [6] H. ATTOUCH, J. BOLTE, AND B. F. SVAITER, *Convergence of descent methods for semi-*
837 *algebraic and tame problems: proximal algorithms, forward-backward splitting, and regu-*
838 *larized Gauss-Seidel methods*, Mathematical Programming, Series A, 137 (2011), pp. 91–
839 124.
840 [7] D. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, 2 ed., 1999.
841 [8] P. BLOMGREN, T. CHAN, P. MULET, AND C. WONG, *Total variation image restoration:*
842 *Numerical methods and extensions*, IEEE International Conference on Image Processing,
843 (1997).
844 [9] J. BOLTE, H. BAUSCHKE, AND M. TEBoulLE, *A descent lemma beyond Lipschitz gradient*
845 *continuity: First-order methods revisited and applications*, Mathematics of Operations
846 Research, 42 (2016).
847 [10] J. BOLTE, A. DANILIDIS, A. LEWIS, AND M. SHIOTA, *Clarke subgradients of stratifiable*
848 *functions*, SIAM Journal on Optimization, 18 (2007), pp. 556–572.
849 [11] J. BOLTE, S. SABACH, AND M. TEBoulLE, *Proximal alternating linearized minimization for*
850 *nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.
851 [12] J. BOLTE, S. SABACH, M. TEBoulLE, AND Y. VAISBOURD, *First order methods beyond con-*

- 852 *velocity and Lipschitz gradient continuity with applications to quadratic inverse problems*,
853 SIAM Journal on Optimization, 28 (2018), pp. 2131–2151.
- 854 [13] S. BONETTI, M. PRATO, AND S. REBEGOLDI, *A block coordinate variable metric linesearch*
855 *based proximal gradient method*, Computational Optimization and Applications, (2018).
- 856 [14] S. BONETTINI, F. PORTA, M. PRATO, S. REBEGOLDI, V. RUGGIERO, AND L. ZANNI, *Recent*
857 *Advances in Variable Metric First-Order Methods*, Springer International Publishing,
858 Cham, 2019, pp. 1–31.
- 859 [15] X. CAI, R. CHAN, M. NIKOLOVA, AND T. ZENG, *A three-stage approach for segmenting*
860 *degraded color images: Smoothing, Lifting and Thresholding (SLaT)*, Journal of Scientific
861 Computing, 72 (2017), pp. 1313–1332.
- 862 [16] X. CAI, R. CHAN, C.-B. SCHÖNLIEB, G. STEIDL, AND T. ZENG, *Linkage between piecewise*
863 *constant Mumford–Shah model and Rudin–Osher–Fatemi model and its virtue in image*
864 *segmentation*, SIAM Journal on Scientific Computing, 41 (2019), pp. B1310–B1340.
- 865 [17] X. CAI, R. CHAN, AND T. ZENG, *A two-stage image segmentation method using a convex*
866 *variant of the Mumford–Shah model and thresholding*, SIAM Journal on Imaging Sciences,
867 6 (2013), pp. 368–390.
- 868 [18] Y. CENSOR AND A. LENT, *Optimization of “log x” entropy over linear equality constraints*,
869 Siam Journal on Control and Optimization, 25 (1987), pp. 921–933.
- 870 [19] A. CHAMBOLLE, D. CREMERS, AND T. POCK, *A convex approach to minimal partitions*, SIAM
871 Journal on Imaging Sciences, 5 (2012), pp. 1113–1158.
- 872 [20] R. CHAN, H. YANG, AND T. ZENG, *A two-stage image segmentation method for blurry images*
873 *with Poisson or multiplicative Gamma noise*, SIAM Journal on Imaging Sciences, 7 (2014),
874 pp. 98–127.
- 875 [21] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT, AND M. BARLAUD, *Deterministic edge-*
876 *preserving regularization in computed imaging*, IEEE Transactions on Image Processing, 6
877 2 (1997), pp. 298–311.
- 878 [22] R. CHARTRAND, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE
879 Signal Processing Letters, 14 (2007), pp. 707 – 710.
- 880 [23] Y. CHEN, S. LEVINE, AND M. RAO, *Variable exponent, linear growth functionals in image*
881 *restoration*, SIAM Journal on Applied Mathematics, 66 (2006), pp. 1383–1406.
- 882 [24] E. CHOUZENOUX, J.-C. PESQUET, AND A. REPETTI, *Variable metric Forward-Backward algo-*
883 *rithm for minimizing the sum of a differentiable function and a convex function*, Journal
884 of Optimization Theory and Applications, 162 (2014), pp. 107–132.
- 885 [25] E. CHOUZENOUX, J.-C. PESQUET, AND A. REPETTI, *A block coordinate variable metric for-*
886 *ward-backward algorithm*, Journal of Global Optimization, (2016), pp. 1–29.
- 887 [26] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing*,
888 Springer New York, New York, NY, 2011, pp. 185–212.
- 889 [27] P. L. COMBETTES AND J.-C. PESQUET, *Fixed point strategies in data science*, IEEE Transac-
890 tions on Signal Processing, 69 (2021), p. 3878–3905.
- 891 [28] P. L. COMBETTES, ĐINH DŨNG, AND BẰNG CÔNG VŨ, *Proximity for sums of composite*
892 *functions*, Journal of Mathematical Analysis and Applications, 380 (2011), pp. 680–688.
- 893 [29] L. CONDAT, *A Primal–Dual splitting method for convex optimization involving lipschitzian,*
894 *proximable and linear composite terms*, Journal of Optimization Theory and Applications,
895 158 (2013).
- 896 [30] M.-C. CORBINEAU, D. KOUAMÉ, E. CHOUZENOUX, J.-Y. TOURNERET, AND J.-C. PESQUET,
897 *Preconditioned P-ULA for joint deconvolution-segmentation of ultrasound images*, IEEE
898 Signal Processing Letters, 26 (2019), pp. 1456–1460.
- 899 [31] R. CORLESS, G. GONNET, D. HARE, D. JEFFREY, AND D. KNUTH, *On the Lambert W*
900 *function*, Advances in Computational Mathematics, 5 (1996), pp. 329–359.
- 901 [32] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for lin-*
902 *ear inverse problems with a sparsity constraints*, Communications on Pure and Applied
903 Mathematics, 57 (2004).
- 904 [33] M. DO AND M. VETTERLI, *Wavelet-based texture retrieval using generalized Gaussian density*
905 *and Kullback-Leibler distance*, IEEE Transactions on Image Processing, 11 (2002), pp. 146–
906 158.
- 907 [34] H. ERDOGAN AND J. FESSLER, *Monotonic algorithms for transmission tomography*, IEEE
908 Transactions on Medical Imaging, 18 (1999), pp. 801–814.
- 909 [35] M. FOARE, N. PUSTELNIK, AND L. CONDAT, *Semi-Linearized proximal alternating minimiza-*
910 *tion for a discrete Mumford–Shah model*, IEEE Transactions on Image Processing, 29
911 (2020), pp. 2176–2189.
- 912 [36] A. GABRIELOV, *Complements of subanalytic sets and existential formulas for analytic func-*
913 *tions*, Inventiones mathematicae, 125 (1996), pp. 1–12.

- 914 [37] D. GHILLI AND K. KUNISCH, *On monotone and primal-dual active set schemes for ℓ_p -type*
915 *problems, $p \in (0, 1]$* , Computational Optimization and Applications, 72 (2019), pp. 45–85.
- 916 [38] M. GRASMAIR, *Well-posedness and convergence rates for sparse regularization with sublinear*
917 *ℓ^q penalty term*, Inverse Problems & Imaging, 3 (2009), pp. 383–387.
- 918 [39] M. GRASMAIR, M. HALTMEIER, AND O. SCHERZER, *Sparse regularization with l_q penalty term*,
919 *Inverse Problems*, 24 (2008), p. 055020.
- 920 [40] J. HERTRICH AND G. STEIDL, *Inertial stochastic palm (ispalm) and applications in machine*
921 *learning*, (2020).
- 922 [41] C. HILDRETH, *A quadratic programming procedure*, Naval Research Logistics Quarterly, 4
923 (1957), pp. 79–85.
- 924 [42] M. HINTERMÜLLER AND T. WU, *Nonconvex TV q -models in image restoration: analysis and a*
925 *Trust-Region regularization-based superlinearly convergent solver*, SIAM Journal on Imag-
926 *ing Sciences*, 6 (2013), pp. 1385–1415.
- 927 [43] D. HUNTER AND K. LANGE, *A Tutorial on MM Algorithms*, The American Statistician, 58
928 (2004), pp. 30–37.
- 929 [44] N. KOMODAKIS AND J.-C. PESQUET, *Playing with duality: An overview of recent primal-dual*
930 *approaches for solving large-scale optimization problems*, IEEE Signal Processing Maga-
931 *zine*, 32 (2015), pp. 31–54.
- 932 [45] K. KURDYKA, *On gradients of functions definable in o -minimal structures*, Annales de l’Institut
933 *Fourier*, 48 (1998), pp. 769–783.
- 934 [46] A. LANZA, S. MORIGI, M. PRAGLIOLA, AND F. SGALLARI, *Space-variant generalised Gaussian*
935 *regularisation for image restoration*, Computer Methods in Biomechanics and Biomedical
936 *Engineering: Imaging and Visualization*, 7 (2018), pp. 1–14.
- 937 [47] H. LE, N. GILLIS, AND P. PATRINOS, *Inertial block proximal methods for non-convex*
938 *non-smooth optimization*, 119 (2020), pp. 5671–5681, <https://proceedings.mlr.press/v119/le20a.html>.
- 939 [48] S. ŁOJASIEWICZ, *Une propriété topologique des sous-ensembles analytiques réels*. Equ. Derivees
940 *partielles*, Paris 1962, Colloques internat. Centre nat. Rech. sci. 117, 87–89 (1963)., 1963.
- 941 [49] S. ŁOJASIEWICZ, *Sur la géométrie semi- et sous-analytique*, Annales de l’Institut Fourier, 43
942 (1993), pp. 1575–1595.
- 943 [50] D. LORENZ, *Convergence rates and source conditions for Tikhonov regularization with sparsity*
944 *constraints*, Journal of Inverse and Ill-posed Problems, 16 (2008), pp. 463–478.
- 945 [51] D. LORENZ AND E. RESMERITA, *Flexible sparse regularization*, Inverse Problems, 33 (2016).
- 946 [52] D. MUMFORD AND J. SHAH, *Optimal approximations by piecewise smooth functions and associ-*
947 *ated variational problems*, Communications on Pure and Applied Mathematics, 42 (1989),
948 pp. 577–685.
- 949 [53] M. NIKOLOVA AND P. TAN, *Alternating proximal gradient descent for nonconvex regularised*
950 *problems with multiconvex coupling terms*. Aug. 2017, <https://hal.archives-ouvertes.fr/hal-01492846>.
- 951 [54] N. OTSU, *A threshold selection method from gray-level histograms*, IEEE Transactions on
952 *Systems, Man, and Cybernetics*, 9 (1979), pp. 62–66.
- 953 [55] B. PASCAL, S. VAITER, N. PUSTELNIK, AND P. ABRY, *Automated data-driven selection of the*
954 *hyperparameters for total-variation based texture segmentation*, Journal of Mathematical
955 *Imaging and Vision*, 63 (2021), pp. 923–952.
- 956 [56] M. PEREYRA, *Proximal Markov chain Monte Carlo algorithms*, Statistics and Computing, 26
957 (2013).
- 958 [57] T. POCK AND S. SABACH, *Inertial proximal alternating linearized minimization (iPALM)*
959 *for nonconvex and nonsmooth problems*, SIAM Journal on Imaging Sciences, 9 (2016),
960 pp. 1756–1787.
- 961 [58] R. RAMLAU AND E. RESMERITA, *Convergence rates for regularization with sparsity constraints*,
962 *Electronic transactions on numerical analysis ETNA*, 37 (2010), pp. 87–104.
- 963 [59] A. REPETTI AND Y. WIAUX, *Variable metric forward-backward algorithm for composite min-*
964 *imization problems*, SIAM Journal on Optimization, 31 (2021), pp. 1215–1241.
- 965 [60] C. ROBERT, V. ELVIRA, N. TAWN, AND C. WU, *Accelerating MCMC algorithms*, Wiley
966 *Interdisciplinary Reviews: Computational Statistics*, 10 (2018).
- 967 [61] R. ROCKAFELLAR AND R. J. WETS, *Variational Analysis*, Springer Verlag, 2004.
- 968 [62] E. D. SCHIFANO, R. L. STRAWDERMAN, AND M. T. WELLS, *Majorization-Minimization al-*
969 *gorithms for nonsmoothly penalized objective functions*, Electronic Journal of Statistics, 4
970 (2010), pp. 1258 – 1299.
- 971 [63] R. TIBSHIRANI, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statis-
972 *tical Society: Series B (Methodological)*, 58 (1996), pp. 267–288.
- 973 [64] J. TOUGERON, *Sur les ensembles semi-analytiques avec conditions gevey au bord*, Annales

- 976 Scientifiques De L Ecole Normale Superieure, 27 (1994), pp. 173–208.
- 977 [65] P. TSENG, *Convergence of a block coordinate descent method for nondifferentiable minimiza-*
978 *tion*, Journal of Optimization Theory and Applications, 109 (2001), pp. 475–494.
- 979 [66] L. VAN DEN DRIES, *Tame Topology and O-minimal Structures*, London Mathematical Society
980 Lecture Note Series, Cambridge University Press, 1998.
- 981 [67] L. VAN DEN DRIES, A. MACINTYRE, AND D. MARKER, *Logarithmic-exponential power series*,
982 Journal of the London Mathematical Society, 56 (1997), pp. 417–434.
- 983 [68] L. VAN DEN DRIES AND P. SPEISSEGER, *The field of reals with multisummable series and*
984 *the exponential function*, Proceedings of The London Mathematical Society, 81 (2000),
985 pp. 513–565.
- 986 [69] B. C. VŪ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*,
987 Advances in Computational Mathematics, 38 (2013), pp. 667–681.
- 988 [70] Z. WANG, A. BOVIK, H. SHEIKH, AND E. SIMONCELLI, *Image quality assessment: from*
989 *error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004),
990 pp. 600–612.
- 991 [71] A. WILKIE, *Model completeness results for expansions of the ordered field of real numbers*
992 *by restricted Pfaffian functions and the exponential function*, Journal of the American
993 Mathematical Society, 9 (1996), pp. 1051–1094.
- 994 [72] J. W. WRENCH, *Concerning two series for the Gamma function*, Mathematics of Computation,
995 22 (1968), pp. 617–626.
- 996 [73] C. ZARZER, *On Tikhonov regularization with non-convex sparsity constraints*, Inverse Prob-
997 lems, 25 (2009), p. 025006.
- 998 [74] N. ZHAO, A. BASARAB, D. KOUAMÉ, AND J.-Y. TOURNERET, *Joint segmentation and decon-*
999 *volution of ultrasound images using a hierarchical Bayesian model based on generalized*
1000 *Gaussian priors*, IEEE Transactions on Image Processing, 25 (2016), pp. 3736–3750.