



**HAL**  
open science

# Space-Variant Image Reconstruction via Cauchy Regularisation: Application to Optical Coherence Tomography

Alin Achim, Luca Calatroni, Serena Morigi, Gabriele Scrivanti

► **To cite this version:**

Alin Achim, Luca Calatroni, Serena Morigi, Gabriele Scrivanti. Space-Variant Image Reconstruction via Cauchy Regularisation: Application to Optical Coherence Tomography. *Signal Processing*, 2022, 20, 10.1016/j.sigpro.2022.108866 . hal-03594202v1

**HAL Id: hal-03594202**

**<https://hal.science/hal-03594202v1>**

Submitted on 2 Mar 2022 (v1), last revised 21 Dec 2022 (v2)

**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.

# Space-Variant Image Reconstruction via Cauchy Regularisation: Application to Optical Coherence Tomography

Alin Achim<sup>\*#</sup>    Luca Calatroni<sup>\*</sup>    Serena Morigi<sup>†</sup>    Gabriele Scrivanti<sup>†\*§</sup>

<sup>†</sup> *Department of Mathematics, University of Bologna, P.zza Porta S. Donato 5, Italy.*

<sup>\*</sup> *Laboratoire I3S, CNRS, UCA, INRIA, 2000 Route des Lucioles, 06903 Sophia-Antipolis, France*

<sup>#</sup> *Visual Information Laboratory, University of Bristol, Bristol, UK.*

<sup>§</sup> *Université Paris-Saclay, Inria, CentraleSupélec, CVN, Gif-sur-Yvette, France.*

---

## Abstract

We propose an adaptive, smooth, non-convex and sparsity-promoting variational model for single-image super-resolution of real murine Optical Coherence Tomography (OCT) data. We follow a sparse-representation approach where sparsity is modelled with respect to a suitable dictionary generated from high-resolution OCT data. To do so, we employ pre-learned dictionaries tailored to model  $\alpha$ -stable statistics in the non-Gaussian case, i.e.  $\alpha < 2$ . The image reconstruction problem renders here particularly challenging due to the high level of noise degradation and to the heterogeneity of the data at hand. As a regulariser, we employ a smooth, non-convex and separable Cauchy-type penalty. To favour adaptivity to heterogeneous image contents, we propose a space-variant modelling by which the local degree of non-convexity encoded patch-wise within the local Cauchy shape parameter is estimated via maximum likelihood. For the solution of the patch-based smooth non-convex optimisation problems, we consider an extension of the cautious Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm where the descent direction is suitably updated depending on the local convexity behaviour of the functional. Our numerical results show that the combination of a space-variant modelling with a fast optimisation strategy improves reconstruction results, maintaining tissue texture and suppressing background noise to a desirable amount at the same time. Furthermore, the proposed optimisation strategy significantly reduces the computational efforts which often represent a major limitation in the analysis of OCT data.

*Keywords:* Sparse representation, Cauchy regularisation, Space-Variant modelling, Non-convex BFGS, Optical Coherence Tomography.

---

## 1. Introduction

At the core of model signal processing methodology sits the concept of sparsity. Working in sparse domains can greatly simplify the computations and produce highly effective and promising results. Sparsity is understood both as sparse signal and image representations (see, e.g., [26, 4, 10, 40]) but also interpreted within a statistical setting and in terms of heavy-tailed probability density functions [36]. As far as the use of sparse representation models for the specific problem of super-resolution is concerned, in their seminal work [42], the authors proposed a dictionary-based approach where patches of a given low-resolution (LR) image were represented in a sparse way with respect to a given over-complete dictionary; from such representation a high-resolution (HR) output was then computed using the estimated coefficients in an efficient way. The proposed approach was motivated by standard results in compressed sensing and highly relied on the use of a suitable choice of sparsity-promoting functionals together with a double dictionary training.

Using redundant representations and sparsity as driving forces for signal and image processing has drawn a lot of research attention in the past decade towards the design of sparsity-regularized variational models [33, 9, 21, 22, 7, 8]. The notion of sparsity can be then described in terms of the minimisation of the non-convex and non-continuous  $\ell_0$  pseudo-norm, i.e. the number of nonzero elements of a given input vector. However, the minimisation of such penalty when combined with, e.g., a quadratic data fidelity makes the composite problem NP-hard [29], so that approximated approaches have to be considered in order to solve the problem. A popular choice

consists in replacing the  $\ell_0$  pseudo-norm with the convex  $\ell_1$  norm [11, 14], which for a certain class of model operators has been shown to provide solutions with the desired sparsity pattern. However, for general operators the choice of such penalty may cause an under-estimation of high amplitude components, an issue which is typically addressed by dropping the convexity constraint and by considering continuous, non-smooth and non-convex functions such as  $\ell_p$  “norms” ( $p < 1$ ), Smoothly Clipped Absolute Deviation, Minimax Concave Penalty (MCP) and Continuous Exact relaxations, see [38] for a unified view on these regularisers with a detailed insights on both the analytical and numerical difficulties encountered when using them. By looking at the analogous notion of sparsity studied, e.g., [27] in a statistical setting, which we prefer to refer to as *statistical sparsity*, a new class of practical sparsity-promoting regularisation terms can be derived. In this context, sparsity has to be defined as a limiting property of a sequence of probability distributions that governs both the rate at which probability accumulates near the origin and the rate at which it decreases elsewhere. In the context of ultrasound and synthetic aperture radar images [1, 3], for instance, the two sparsity notions have been related by observing that wavelet coefficients of such images can be accurately characterised statistically by the family of heavy-tailed alpha-stable distributions [36, 30]. Subsequently, Bayesian estimators recovering the signal component of image coefficients in an optimal fashion were designed. One should note that unlike other distributions employed for data modelling, which are generically empirically selected, the use of alpha-stable densities is rigorously justified by the generalised central limit theorem [36]. More recently, a variational non-convex, but smooth penalty function derived from the Cauchy distribution (a particular case of the alpha-stable family with  $\alpha = 1$ ) was similarly employed as a sparsity-promoting penalty for a number of inverse imaging problems [20, 18, 19]. The smoothness of such penalty is particularly interesting from an optimisation point of view as it allows in principle a handy numerical treatment.

Following [20, 18, 19], we consider in the following a sparsity-based approach relying on the use of a Cauchy penalty for the high-resolution reconstruction of biomedical data using redundant and dictionary-based representations. As an exemplar biomedical application, we focus on real Optical Coherence Tomography (OCT) images, with the intent of showing the benefits of applying super-resolution before performing any further image analysis. OCT is an *in vivo* non-invasive imaging technique based on low-coherence interferometry that allows to detect ophthalmic structures at micrometer resolution. OCT images show sections of the multiple layers of the retinal tissue as well as the inner eye region (*vitreous*), hence they are particularly suited for the detection of anomalies and deformations in the eyes as well as in the follow-up of ophthalmic diseases in early and later stages such as Multiple Sclerosis, Diabetes type 1, Alzheimer’s disease, Parkinson disease, or Glaucoma, see, e.g., [45] for a medical review. However, the poor spatial resolution and the multiplicative nature of the (speckle) noise observed in OCT data often limit the possibility of an accurate image analysis, which makes the use of both super-resolution (SR) and denoising/despeckling imaging techniques crucial for the subsequent image analysis, often based on accurate (and often manual) layer segmentation [43, 41]. A reliable reconstruction of OCT images should preserve the essential diagnostic information in the OCT images which is confined to the diagnostically significant (DS) regions of the OCT images characterised by the layered dense structure, while denoising the homogeneous regions consisting of the bottom part of the image (choroid and the sclera) named as diagnostically non-significant (DN) regions [13] which occupy the majority of the pixels in the OCT images. A tailored reconstruction method for the analysis of OCT data should thus adapt to such data variability by means, e.g., of a space-adaptive regularisation. For example, in [15] the authors proposed a sparsity-based simultaneous denoising and interpolation of OCT images, while in [13] the OCT images were super-resolved using a weighted joint sparse representation method to adaptively reconstruct the DS regions while denoising the DN ones. More recently, a generative adversarial network-based approach to simultaneously denoise and super-resolve OCT images was introduced in [16]. However, the utility of deep learning approaches in medical imaging is still a challenge due to the scarcity of large databases of paired LR-HR images.

Our contribution consists in proposing a mathematical approach for the super-resolution of OCT data using the framework of sparse-representation in terms of over-complete dictionaries [42] and a Cauchy-based regulariser as a sparsity-promoting term. Our proposal extends the one preliminarily studied in [37] in several ways comprising:

- (i) a new sparse representation-based reconstruction approach in a super-resolution framework;

- (ii) an adaptive (space-variant) Cauchy-type regularisation model along with an automatic parameter estimation strategy;
- (iii) a tailored quasi-Newton numerical optimisation scheme computing effective model solutions.

We first design a reconstruction approach for OCT images based on their sparse representation with respect to a fixed high resolution (HR) dictionary where the dictionary-construction procedure is performed as a preprocessing step on a small set of properly denoised HR images. Compared to classical approaches, see, e.g., [42], the proposed approach is rather flexible as on one side, it does not require the learning of a low resolution (LR) dictionary, while on the other it allows to reconstruct different LR data, regardless of their initial resolution, once the HR dimension is given. Furthermore, our reference dictionary is built on HR cleaned data, which introduces an intrinsic denoising effect within the reconstruction process.

As a second contribution, we propose an adaptive strategy estimating the variability of the observed imaging data, which we exploit to define a space-variant Cauchy regularisation term whose local degree of regularisation weight and non-convexity is adapted to the local image content. As the proposed estimation strategy relies entirely on the use of the given LR data, the proposed regularisation approach is totally *parameter-free*. The proposed tuning calibrates locally the non-convex behaviour of the regularisation, adapting it to the different regions in the image.

As far as the numerical solution of the resulting space-variant model is concerned, we exploit the overall smoothness of the composite problem and propose an efficient quasi-Newton scheme relying on a modification of the cautious Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm introduced in [25]. In [20, 37], the Cauchy penalty was dealt with within a first-order proximal gradient scheme by exploiting a handy computation of the proximal operator associated to such penalty. Although it was shown to achieve state-of-the-art performance in applications such as SAR image reconstruction [18], medical ultrasound image analysis [19], and MIMO signal detection [20], the use of the Cauchy penalty through its proximal operator does not fully exploit the smoothness of such penalty. Motivated by this observation, we thus introduce an efficient Non-Convex Cautious BFGS (NNC-BFGS) algorithm adapted to solve the composite Cauchy-based reconstruction problem in a more efficient way.

With a slight abuse of terminology, we will refer from now on to our reconstruction approach as *super-resolution* (SR), for consistency with the standard dictionary-based SR approach proposed in [42]. Nevertheless, we highlight two main points for which our model differs from standard dictionary-based ones: firstly, ours relies upon a single HR dictionary and not on a LR-HR pair; secondly, our reconstruction process favours not only a faithful HR reproduction of the DS regions, but also an intrinsic denoising process.

The rest of the paper is organised as follows. In Section 2, the SR OCT problem is formulated: Section 2.1 briefly introduces the Cauchy-based variational model for single patch reconstruction, which is applied in the overall SR framework described in Section 2.2. Sections 3, 4 and 5 describe in detail the main ingredients of the proposed algorithm, respectively the role of the Cauchy-based penalty, the definition and the estimation of the space-variant  $\gamma$ -map and the smooth optimisation method designed to solve the problem. In Section 6 the proposed method is validated on several OCT data and, finally, in Section 7 some conclusions are reported.

## 2. OCT SR via sparse and adaptive representation

For  $r_h, c_h, r_l, c_l \in \mathbb{R}_+$  with  $r_l \leq r_h$  and  $c_l \leq c_h$ , the task of recovering an HR OCT image  $X \in \mathbb{R}^{r_h \times c_h}$  from a noisy, and blurred LR input  $Y \in \mathbb{R}^{r_l \times c_l}$  can be modelled mathematically as an inverse image reconstruction problem whose ill-posedness can be overcome by representing  $X$  in a sparse way with respect to a given (over-complete) dictionary [42] and by introducing the following two constraints:

1. *Reconstruction Constraint*: the input LR image  $Y$  is linked to the desired HR image  $X$  via the image formation model

$$Y = \mathcal{N}(\mathcal{S}_q(\mathcal{K}(X))) + \eta \quad (\text{C1})$$

where the operator  $\mathcal{K} : \mathbb{R}^{r_h \times c_h} \rightarrow \mathbb{R}^{r_l \times c_l}$  is the convolution (blur) operator corresponding to the point spread function (PSF)  $\kappa \in \mathbb{R}^{s_1 \times s_2}$  of the OCT acquisition system, the matrix

$\mathcal{S}_q : \mathbb{R}^{r_h \times c_h} \rightarrow \mathbb{R}^{r_h/q \times c_h/q}$  is a downsampling operator defined in terms of a factor  $q \in \mathbb{N}^*$  which maps the image  $\mathcal{K}(X) = \kappa * X \in \mathbb{R}^{r_h \times c_h}$  into a coarser grid by averaging/interpolation. As any imaging technique based on detection of coherent waves, OCT images are subject to the presence of speckle noise [35]. Consequently,  $\mathcal{N}(\cdot)$  is used here to denote a multiplicative (speckle) noise degradation process, while  $\eta \in \mathbb{R}^{r_h/q \times c_h/q}$  denotes an additive component representing white Gaussian measurement noise.

2. *Sparsity Constraint*: We further assume that every (square) patch  $x$  extracted from  $X \in \mathbb{R}^{r_h \times c_h}$  can be represented as a sparse linear combination of  $n_d$  atoms of a given over-complete dictionary  $\mathcal{D} \in \mathbb{R}^{n_p \times n_d}$  which has been previously learned from HR training images. Using a vectorised notation for  $x \in \mathbb{R}^{n_p}$ , such assumption can thus be formulated as:

$$x \approx \mathcal{D}a \quad \text{for some } a \in \mathbb{R}^{n_d} \quad \text{with } \|a\|_0 \ll n_d, \quad (\text{C2})$$

where by  $\|a\|_0 := \#\{a_i : a_i \neq 0\}$  we denote the  $\ell_0$  pseudo-norm of the vector  $a$ .

### 2.1. Patch reconstruction via adaptive space-variant modelling

As discussed in the introduction, OCT data show very heterogeneous contents: different regions of the images considered enclose in fact varying levels of texture information, which translate into diverse sparsity patterns of the  $n_d$  atoms ( $n_d > n_p$ ) with respect to given dictionary  $\mathcal{D} \in \mathbb{R}^{n_p \times n_d}$ . We then combine (C1) with (C2) to define the local relation between each vectorised square patch  $y \in \mathbb{R}^{n_p/q^2}$  extracted from the LR observed image  $Y$  and the corresponding coefficient vector  $a_y$ , thus finally writing the patch-based sparse representation problem:

$$\text{find } a_y \in \mathbb{R}^{n_d} \quad \text{such that } y = \mathcal{N}(\mathcal{S}_q(\mathcal{K}(\mathcal{D}a_y))) + \eta \quad \text{with } \|a_y\|_0 \ll n_d. \quad (M_y)$$

Solving the reconstruction problem under model  $(M_y)$  is challenging since:

- **[Ill-Posedness]** it is an inherently ill-posed problem;
- **[NP-hardness]** it is NP-hard, due to the presence of the  $\ell_0$  pseudo-norm;
- **[Heterogeneity]** it is patch-dependent, due to the heterogeneous (possibly very different) content of each image patch.

To address these challenges, we consider in the following a variational space-variant and sparsity-inducing approach. In particular, we consider a penalised variational problem for the retrieval of the coefficients  $a_y$  where the shape parameter of the penalty considered is tuned depending on the content of each patch  $y$ . To do so, we define a space-variant prior information function, called in the following  $\gamma$ -map, which assigns to each image patch  $y$  of the LR image a value  $\gamma_y > 0$  based on a local estimation of the sparsity level of  $y$ . The details about the choice of this particular function will be outlined in Section 4. Starting from  $(M_y)$ , we thus consider the following variational optimisation problem

$$\text{find } a_y^* \in \underset{a \in \mathbb{R}^{n_d}}{\text{argmin}} \{f(a) := f_1(a) + f_2(a; \gamma_y)\} \quad (P_y)$$

$$f_1(a) := \frac{1}{2} \|y - \mathcal{S}_q \mathcal{K} \mathcal{D}(a)\|_2^2, \quad f_2(a; \gamma_y) := \sum_{i=1}^{n_d} \phi(a_i; \gamma_y),$$

where, for any LR patch  $y$ ,  $a_y^*$  is an approximation of the sparse coefficient vector representing the HR patch  $x \in \mathbb{R}^{n_p}$  in terms of the over-complete dictionary  $\mathcal{D}$ .

A quadratic data fidelity term  $f_1(\cdot)$  is used to remove AWG noise, while the regularisation term  $f_2(\cdot; \gamma_y)$  is defined in terms of the separable function  $\phi(\cdot; \gamma_y)$  which we choose to be the smooth, non-convex Cauchy penalty of shape parameter  $\gamma_y > 0$  previously employed, e.g., in [20, 37]. The patch-dependent parameter  $\gamma_y$  plays a crucial twofold role in our model. Firstly, it plays the role of the regularisation parameter, balancing the contribution of the fidelity term against the Cauchy penalty. Secondly, it modulates the non-convexity and hence the sparsity-promoting property of the penalty. For this reason, such parameter will be referred to as *convexity parameter*.

In comparison with our previous work [37], the super-resolution model  $(P_y)$  extends the Cauchy-based regularisation strategy to the space-variant case which allows to deal effectively with the heterogeneous image contents.

## 2.2. Image reconstruction from patches

We synthesise in Algorithm 1 the main steps of the proposed adaptive OCT-SR method. The matrices  $D$  and  $S_q$  denote suitably resized matrices computed by discretising the operators  $\mathcal{D}$  and  $\mathcal{S}_q$ , respectively. The matrix  $K$  represents the blurring operator corresponding to the PSF of the OCT scan. Such operator can be either considered to be known (assuming, for instance, a Gaussian PSF  $\kappa$ ) or estimated as described in [28].

Algorithm 1 relies only on the HR dictionary  $\mathcal{D}$ , thus allowing greater flexibility in the choice of the scaling factor  $q$  which can be freely varied without recomputing the LR dictionary  $\mathcal{D}$  as required in standard approaches [42]. To build the HR dictionary  $\mathcal{D}$ , we follow [32] where general  $\alpha$ -Stable distributions are used as prior Probability Density Functions (PDF). By setting  $\alpha = 1$ , we thus fix the underlying distribution to be Cauchy. Once we have extracted from  $Y$  the patches  $y$ -s (in raster scan order with overlapping borders of dimension 1 pixel), Algorithm 1 consists in the computation of the solution  $a_y^*$  to problem  $(P_y)$ . We will generically denote by NNC-BFGS the resolution method chosen to address  $(P_y)$ , of which a detailed description is postponed to Section 5.

The reassembling of a HR image  $X$  from the reconstructed patches is then simply obtained following a patch-overlapping strategy. Denoting by  $\Omega_l \subset \mathbb{R}^2$  and  $\Omega_h \subset \mathbb{R}^2$  the LR and HR image spaces, we consider a LR square patch  $y$  having its upper-left corner in position  $(i_l, j_l) \in \Omega_l$ . The position  $(i_h, j_h) \in \Omega_h$  of the upper-left corner of the corresponding HR patch  $x$  is then identified by means of the following projection map  $\Pi_q$ :

$$\begin{aligned} \Pi_q : \Omega_l &\rightarrow \Omega_h \\ (i_l, j_l) &\mapsto ((i_l - 1)q + 1, (j_l - 1)q + 1). \end{aligned}$$

As we are extracting overlapping patches, the reassembling procedure based on  $\Pi_q$  generates a set of maximum  $n_p$  values insisting on the same HR pixel  $(i_h, j_h)$ . A common method is to average these pixel values for obtaining the HR image value in  $\Omega_h$ . However, in order to avoid the smoothing effect generated by averaging, we propose instead to stack all these values and take the median one.

---

### Algorithm 1 Adaptive OCT-SR

---

**Input:**  $Y, D, S_q$

**Output:**  $X^*$  % HR OCT image

Estimate PSF as in [28] to compute  $K$

Estimate the  $\gamma$ -map (see Section 4)

Extract overlapping patches of size  $\frac{\sqrt{n_p}}{q} \times \frac{\sqrt{n_p}}{q}$

**For** each patch  $y$

Compute  $a_y^*$  by solving  $(P_y)$  by NNC-BFGS (see Section 5)

Generate the HR patch  $x_y = D a_y^*$

**end**

Stack the  $x_y$  patches according to  $\Pi_q$

Compute median values of all overlapping pixels

Collect all the median values in  $X^*$

---

## 3. Sparsity-promoting Cauchy penalty

In this section we outline some important properties of the scalar Cauchy penalty functions  $\phi(\cdot; \gamma_y)$  which contribute to the solvability of  $(P_y)$ .

From a statistical point of view, the Cauchy distribution belongs to the family of  $\alpha$ -stable distributions, which are heavy-tailed distributions frequently exploited in many applications ranging from actuarial, financial sciences, risk management and tomographic imaging see, for example [39, 43, 2]. More precisely, a Cauchy distribution is an  $\alpha$ -stable distribution with  $\alpha = 1$ . Differently from other distributions in the family, the probability density function of Cauchy distribution, centred at the origin, can be expressed in closed-form in terms of a positive shape parameter  $\gamma > 0$ . Its expression reads:

$$p(t; \gamma) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + t^2}, \quad t \in \mathbb{R}. \quad (1)$$

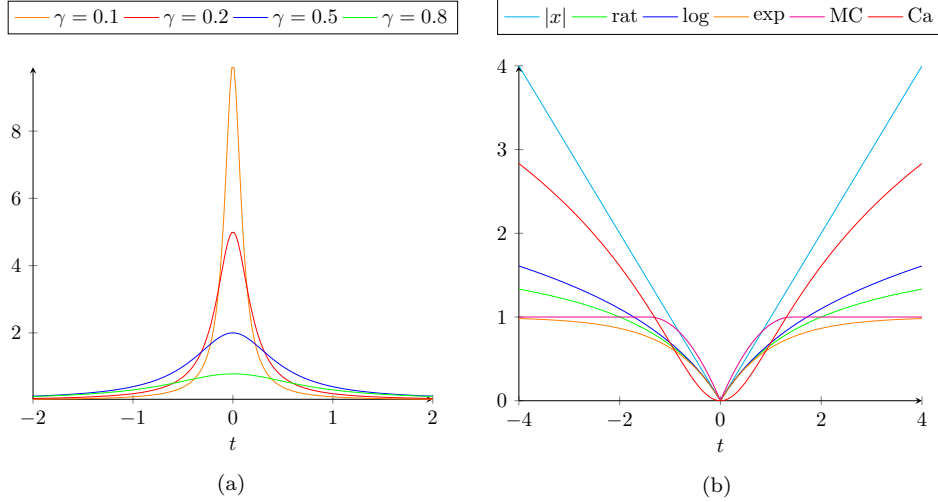


Figure 1: (a) Cauchy PDF for varying  $\gamma > 0$ . (b) Plots of different penalty functions relaxations of the  $\ell_0$  pseudo-norm.

As illustrated in Fig. 1a the value of  $\gamma$  controls the spread of the Cauchy distribution, playing thus the very same role as the Gaussian variance: the smaller  $\gamma$ , the narrower and more peaked the shape of the distribution. The Cauchy penalty can be defined by taking the negative log-likelihood of the corresponding Cauchy PDF. For  $t \in \mathbb{R}$ , such penalty thus reads:

$$\phi(t; \gamma) = -\log\left(\frac{\gamma}{\gamma^2 + t^2}\right) = \log\left(\frac{\gamma^2 + t^2}{\gamma}\right), \quad (2)$$

with first and second derivatives given by:

$$\phi'(t; \gamma) = \frac{2t}{t^2 + \gamma^2}, \quad \phi''(t; \gamma) = \frac{2(\gamma^2 - t^2)}{(t^2 + \gamma^2)^2}. \quad (3)$$

Fig. 1b illustrates a comparison between the Cauchy penalty (labelled as **Ca**), the  $\ell_1$  norm and other popular penalties typically used as alternatives to  $\ell_0$  pseudo-norm, compare e.g. with [24, 38]. Note that the function  $\phi(\cdot; \gamma)$  in (2) is non-convex except for a small and limited interval around the origin, and increases unbounded at a logarithmic rate. This latter property, formulated in Prop.1, is fundamental both to prove the existence of solutions for problem  $(P_y)$  with the penalty in (2), as stated by the Prop.2, and for the fine-tuning of an efficient algorithmic solution, as described in Section 5.

**Proposition 1.** *The function  $\phi(x; \gamma)$  defined in (2) with  $\gamma > 0$ , is twice continuously differentiable, coercive, bounded from below by zero, and non-convex except for the interval  $-\gamma \leq x \leq \gamma$ .*

As a consequence of Proposition 1 the Hessian matrix  $\mathcal{H}_f \in \mathbb{R}^{n_d \times n_d}$  of the function  $f(a)$  in  $(P_y)$  is indefinite and reads as follows

$$\mathcal{H}_f(a) = (S_q K D)^T (S_q K D) + \text{diag}\left(\phi''(a_i; \gamma)\right)_{i=1, \dots, n_d}, \quad (4)$$

where  $\phi''(\cdot)$  is defined as in (3), and  $S_q K D \in \mathbb{R}^{\frac{n_p}{q^2} \times n_d}$ . The existence of a global minimiser for problem  $(P_y)$  can be proved by standard arguments.

**Proposition 2.** *Let  $y \in \mathbb{R}^{n_p/q^2}$ ,  $\gamma_y > 0$  and  $\phi(\cdot; \gamma_y) : \mathbb{R}_+ \rightarrow \mathbb{R}$  be defined in (2). Then, the function  $f$  in  $(P_y)$  is non-convex, proper, continuous, bounded from below by zero, and coercive, hence the problem  $(P_y)$  admits a global minimiser.*

As an alternative to the Cauchy penalty in (2), the convex  $\ell_1$  or the non-convex Minimax Concave Penalty (MCP) penalties can be considered as different examples sparsity-promoting

regularisers, similarly as what has been previously done in [37]. Thanks to its intrinsically tunable degree of non-convexity, MCP is indeed advised to be among the most flexible and effective non-smooth non-convex penalties [23]. However, its lack of smoothness forces to the use of non-smooth optimisation methods such as proximal point algorithms [12, 17] to deal with it from a numerical point of view. Analogous considerations hold for  $\ell_1$  regularisation. On the other hand, according to Proposition 1, the function  $\phi$  is twice continuously differentiable in  $\mathbb{R}$ , so it is the cost function in (2).

**Remark 1.** *When the Cauchy penalty (2) is replaced by the MCP penalty, then Proposition 2 does not hold anymore, as the overall objective function is not coercive any longer as the down-sampling operator  $S_q$  has a nontrivial kernel and the penalty is bounded from above by a constant and from below by zero. Note also that the linear operator  $S_q K D \in \mathbb{R}^{n_p/q^2 \times n_d}$ , with  $n_p/q^2 < n_d$ , cannot be full column rank for over-complete dictionaries, and the penalty is additive separable, thus we cannot resort to convex/non-convex strategies to derive conditions on the parameter  $\gamma$  guaranteeing that the associated functional in  $(P_y)$  is convex as done, e.g., in [23].*

#### 4. Design of the space-variant $\gamma$ -map

OCT image content is very heterogeneous. It is then desirable to design a space-variant sparsity map so as to obtain different regularisation behaviours adapting to the different patches of the observed image. The definition of the Cauchy regulariser  $\phi(\cdot; \gamma_y)$  in  $(P_y)$  for each patch  $y$  can thus take advantage of a patch-dependent parameter  $\gamma_y > 0$  adapted to the local representation of the desired solution.

With the intent of defining local sparsity-promoting information, we thus define the  $\gamma$ -map as the function  $\gamma : \Omega_l \rightarrow \mathbb{R}_+$  which associates to every pixel  $(i, j)$  in the LR image a value  $\gamma_{i,j} > 0$ . For its construction, we thus consider for each pixel  $(i, j) \in \Omega_l$  a neighbourhood  $\mathcal{N}_{i,j} \subset \Omega_l$  of fixed size (which results in a patch  $\tilde{y} = \{\tilde{y}_{\ell,\kappa}\}_{(\ell,\kappa) \in \mathcal{N}_{i,j}}$  that is bigger than  $y$ ) to determine the  $\gamma_{i,j}$  based on an estimation strategy taking advantage of local intensity information. For the given OCT data, we consider a Maximum Likelihood (ML) estimator of the parameter  $\gamma$  which can be written as

$$\gamma_{i,j} \in \arg \max_{\gamma} \sum_{(\ell,\kappa) \in \mathcal{N}_{i,j}} \log \left( \int_{-\infty}^{+\infty} p(\tilde{y}_{\ell,\kappa} - x; \gamma) e^{-\frac{x^2}{2\sigma^2}} dx \right) \quad (5)$$

where  $p$  is the PDF in (1),  $P_\sigma(x) = e^{-\frac{x^2}{2\sigma^2}}$  models Gaussian noise with given standard deviation  $\sigma > 0$ , see [2] for details. The integral in expression (5) can be evaluated using a simple MonteCarlo approximation.

Once the  $\gamma_{i,j}$  values have been estimated for each pixel  $(i, j) \in \Omega_l$  and normalised in  $(0, 1]$ , then the whole  $\gamma$ -map is rescaled into a fixed range  $[\underline{\gamma}, \bar{\gamma}]$  where  $0 < \underline{\gamma}, \bar{\gamma}$  whose extreme values represent the maximally sparse and the minimally sparse behaviour induced by  $\gamma$  in the penalty  $\phi$  in (2). This allows to better enhance the difference between its convex and its nonconvex behaviour, and properly balance the regularisation term. To perform such rescaling we apply the nonlinear sigmoid function

$$s : [0, 1] \rightarrow [0.02, 1] \\ t \mapsto 0.5 + 0.5 \tanh(6(t - 0.3)),$$

which is the strictly increasing function illustrated in Fig. 2(left).

Finally, the value for  $\gamma_y$  appearing in  $(P_y)$  is computed by averaging all the values  $\gamma_{i,j}$  as

$$\gamma_y = \frac{1}{\sqrt{n_p}} \sum_{(i,j) \in \mathcal{I}_y} \gamma_{i,j}, \quad (6)$$

where  $\mathcal{I}_y \subset \Omega_l$  is the subset containing the coordinates of the pixels in the patch  $y$ .

For the sake of illustration, we report in Fig. 2 (right) the  $\gamma$ -map associated to the LR OCT image in the middle, with values  $\gamma_{i,j} \in [0.1, 1]$  for all  $i, j$ .



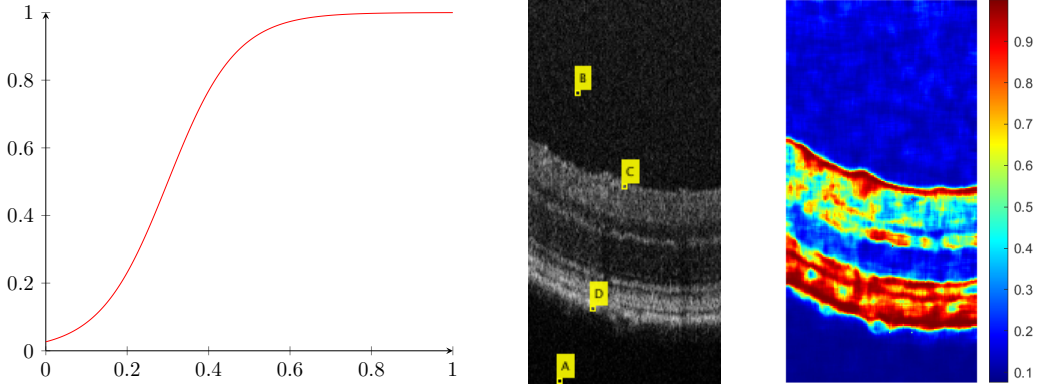


Figure 2: Rescaling sigmoid function (left); LR OCT image with four different labelled patches (centre); corresponding  $\gamma$ -map (right).

To better highlight the link between the  $\gamma$ -value and the local morphological characteristic of the image, we extracted four patches from different regions of the LR OCT image: one in the lower region (patch A), one in the vitreous (patch B), two in the layers (patches C and D), see Fig. 2 (middle). The benefits of using a space-variant regularisation strategy are now illustrated in Fig.3 (first row) for both the OCT image shown in Fig. 2 and for another OCT image in Fig.3 (second row) with more heterogeneous content in the upper region representing suspended particles in the vitreous possibly caused by an inflammation of the eye. In Fig. 3 (left panel) the reconstructed OCT images are obtained by solving the space-variant problem ( $P_y$ ) where the value  $\gamma_y$  is extracted from the corresponding  $\gamma$ -maps in Fig. 3(left). For comparisons, the right panel of Fig. 3, from left to right, shows the reconstructed OCT images obtained by applying the reconstruction problem ( $P_y$ ) with several  $\gamma$  values, fixed for each patch, in the range  $\gamma \in \{0.1, 0.2, 0.5, 0.9\}$ , respectively. As the parameter  $\gamma$  increases, the textured information from the LR OCT image is preserved better and better, as well as the undesired noise in the background. By applying the proposed space-variant approach, textured details are preserved in the DS layers region, while the noise is suppressed in the background, DN regions.

For each selected patch, we report in Table 1 the PSNR values obtained by comparing the solutions of ( $P_y$ ) for the different choice of constant  $\gamma$  and space-variant  $\gamma_y$ . The PSNR values have been computed with respect to the corresponding patches extracted from the Ground Truth HR image. We observe that the automatic space-variant procedure allows to obtain high-quality reconstructions whose PSNR values are comparable to the ones obtained by manually tuning the  $\gamma$  parameter, thus reducing the computational efforts without affecting (if not improving) the reconstruction quality.

# Patch \ $\gamma$	0.1	0.2	0.5	0.9	SV $\gamma$
A (background)	<b>38.053</b>	38.041	37.886	37.584	<b>38.055</b> (0.08)
B (vitreous)	24.157	<b>24.172</b>	24.137	24.046	24.162 (0.14)
C (layers)	25.642	26.183	27.295	<b>27.504</b>	<b>27.525</b> (0.86)
D (layers)	17.264	18.651	22.840	<b>24.957</b>	<b>25.072</b> (0.99)

Table 1: PSNR values for SR of selected patches by solving ( $P_y$ ) with constant  $\gamma \in \{0.1, 0.2, 0.5, 0.9\}$  and space-variant  $\gamma_y$  where the estimated  $\gamma_y$  values are reported in brackets. Best PSNR values are boldfaced.

## 5. Smooth, non-convex optimisation via cautious BFGS

From an optimisation point of view, the patch-dependent problem ( $P_y$ ) is small dimensional and smooth, although possibly non-convex. The gradient  $\nabla f(\cdot)$  of the cost function  $f$  in ( $P_y$ ) is Lipschitz continuous with constant  $L > 0$  that could be estimated by triangle inequality as  $L = L_{f_1} + L_{f_2}$  where  $L_{f_1}$  and  $L_{f_2}$  are, respectively, the Lipschitz constants of the quadratic

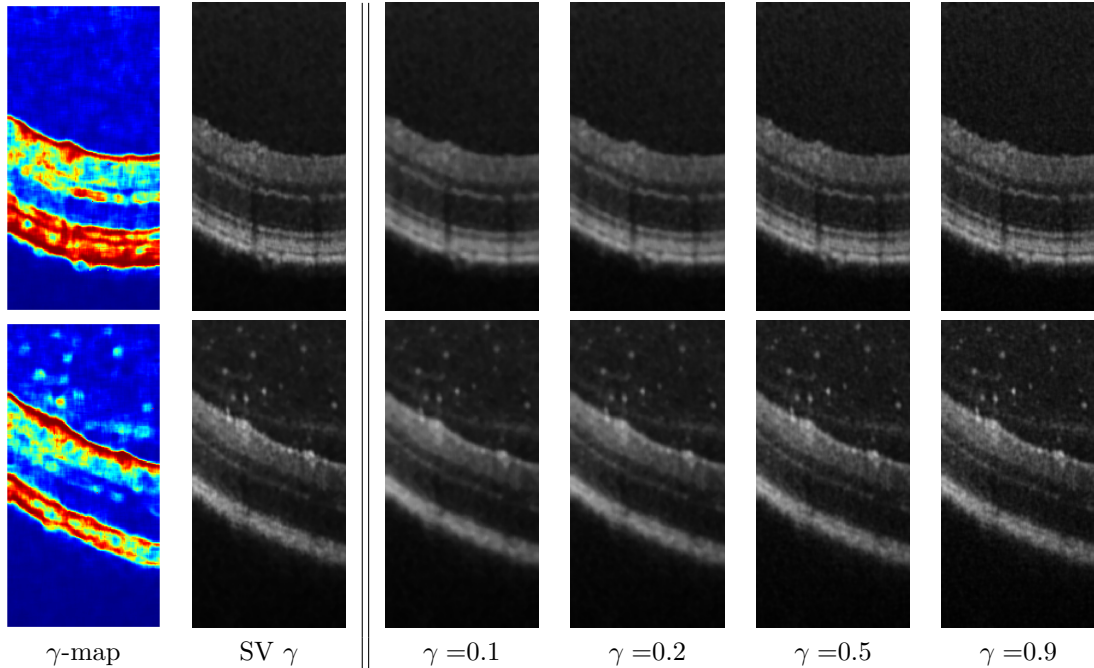


Figure 3: Comparisons between the proposed space-variant (SV) reconstruction approach (left) and the one with fixed values of  $\gamma$  (right) on the images OCT1 (first row) and OCT2 (second row), obtained with an upscaling factor  $q = 4$ .

data-term and Cauchy penalty terms in  $(P_y)$ , that is

$$L_{f_1} = \|(S_q H D)^T S_q H D\|_2^2, \quad L_{f_2} = \frac{2}{\gamma(x)} \leq \frac{2}{\underline{\gamma}} < +\infty. \quad (7)$$

Problem  $(P_y)$  can thus be solved efficiently by means of standard smooth optimisation algorithms, such as, e.g., Newton-type methods, without resorting to first-order proximal gradient methods as previously done in, e.g., [41]. In particular, the small dimension of the sub-problem at hand allows to incorporate Hessian-type information by means of the BFGS algorithm whose convergence, in the case of a convex function  $f(\cdot)$ , is super-linear.

### 5.1. A cautious BFGS update

The BFGS is a well-known quasi-Newton method that allows to solve smooth unconstrained optimisation problems. Its starting point consists in considering the following quadratic model for the objective function  $f$  to minimise at the iteration  $k \geq 1$ :

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p, \quad (8)$$

where  $B_k$  is a symmetric positive definite matrix approximating the Hessian of  $f$ . The method proceeds by identifying the descent direction as the minimum point of (8) according to

$$p_k = -(B_k)^{-1} \nabla f_k, \quad (9)$$

$$x_{k+1} = x_k + \lambda_k p_k, \quad (10)$$

where  $\lambda_k > 0$  is the step-size parameter. Let  $B_0$  be any given symmetric positive-definite matrix, then the updated matrix  $B_k$  at every iteration can be written by using the recursion formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{r_k r_k^T}{r_k^T s_k}, \quad (11)$$

where  $s_k := x_{k+1} - x_k$ ,  $r_k := g_{k+1} - g_k$ , and  $g_k := \nabla f(x_k)$ . Although the local convergence of this method is well understood [31], its global behaviour in a non-convex setting is still an open question.

In [25] the authors propose a variation of BFGS based on a *cautious update* of the matrix  $B_k$  (CBFGS) for which global and superlinear convergence is proved, under the assumption of Wolfe-type and Armijo-type line-search strategies and Lipschitz continuous gradient continuity of the objective function. For a suitable choice of parameters  $\varepsilon, \alpha > 0$ , the proposed cautious update guarantees that the matrix  $B_{k+1}$  generated is symmetric and positive definite for all  $k$ , and takes the following form

$$B_{k+1} = \begin{cases} B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{r_k r_k^T}{r_k^T s_k} & \text{if } \frac{r_k^T s_k}{\|s_k\|^2} > \varepsilon \|g_k\|^\alpha, \\ B_k & \text{otherwise.} \end{cases} \quad (12)$$

The choice of the parameters  $\varepsilon$  and  $\alpha$  in (12) clearly affects the practical application of the update rules. In [25] the authors suggest to choose a fixed  $\varepsilon$  and to either vary  $\alpha$  according to the following criterion

$$\alpha = \begin{cases} 0.01 & \text{if } \|g_k\| > 1, \\ 3 & \text{if } \|g_k\| \leq 1, \end{cases} \quad (\text{CBFGS}^1)$$

or to set  $\alpha = 1$  (CBFGS<sup>2</sup>).

## 5.2. Normalised NonConvex BFGS

Problem  $(P_y)$  is convex whenever the penalty  $\phi(\cdot; \gamma)$  is convex which, by Proposition 1 corresponds to the range  $[-\gamma, \gamma]$ . Outside this range,  $(P_y)$  is non-convex, hence ordinary BFGS method is not guaranteed to converge. Taking advantage of the information encoded in the  $\gamma$ -map, we propose an update of the descent direction in order to deal *cautiously* with non-convex subproblems. This can be done by setting a thresholding parameter  $\gamma_{th} \in [\underline{\gamma}, \bar{\gamma}]$ , beneath which the problem is treated cautiously due to local high non-convexity. Instead of following the cautious strategy detailed above, we introduce in the following a cautious choice of the descent direction update, which reads:

$$\tilde{p}_k = -H_k g_k, \quad (13)$$

$$p_k = \begin{cases} \tilde{p}_k & \text{if } \gamma > \gamma_{th} \\ \frac{\tilde{p}_k}{\|\tilde{p}_k\|} & \text{otherwise,} \end{cases} \quad (14)$$

where  $H_k$  is the inverse of  $B_k$ , which can be updated as follows

$$H_{k+1} = (I - \rho_k s_k r_k^T) H_k (I - \rho_k r_k s_k^T) + \rho_k s_k s_k^T, \quad \rho_k = 1/(r_k^T s_k). \quad (15)$$

Our proposal (13)-(15) simply consists in normalising the descent direction so as to allow the inexact linesearch to look for an overall appropriate step-size  $\lambda_k$  in (10). For small values  $\gamma \leq \gamma_{th}$  the Cauchy penalty  $\phi(a; \gamma)$  is non-convex and its second derivative has a great magnitude that affects the nonlinear part of the Hessian matrix in (4). Thus the scaling on  $p_k$  prevents the approximation  $H_k$  from becoming too large, thus helping in the overall stabilisation of the regularisation process.

As a line-search strategy, we consider here the *strong Wolfe conditions*, which require that at each  $k \geq 1$  the stepsize  $\lambda_k$  to satisfy a *sufficient decrease condition* (W1) and a *strong curvature condition* (W2) for some constants  $0 < \sigma_1 < \sigma_2 < 1$ :

$$f(x_k + \lambda_k p_k) \leq f(x_k) + \sigma_1 \lambda_k \nabla f_k^T p_k, \quad (\text{W1})$$

$$|\nabla f(x_k + \lambda_k p_k)^T p_k| \leq \sigma_2 |\nabla f_k^T p_k|. \quad (\text{W2})$$

It is possible, upon standard arguments, to prove that the proposed Normalised Non-Convex BFGS (NNC-BFGS) update endowed with the Strong Wolfe line search preserves the positive semi-definiteness.

**Proposition 3.** *Let the NNC-BFGS method with strong Wolfe conditions be applied to a continuously differentiable function  $f$  that is bounded from below. Assume  $H_0$  is symmetric and positive definite. Then, for all  $k \geq 1$ ,  $H_{k+1}$  is positive definite.*

*Proof.* We first notice from (15) that the thesis holds if and only if  $\langle r_k, s_k \rangle = \langle g_{k+1} - g_k, x_{k+1} - x_k \rangle > 0$ . By induction, let us assume that  $H_k$  is positive definite. Since  $s_k = x_{k+1} - x_k = \lambda_k p_k$  with  $p_k$  defined as in (13)- (14), we have:

$$\begin{aligned}
\langle g_{k+1} - g_k, x_{k+1} - x_k \rangle &= \lambda_k \langle g_{k+1} - g_k, p_k \rangle \\
&= \lambda_k \langle g_{k+1}, p_k \rangle - \lambda_k \langle g_k, p_k \rangle \\
&\geq \lambda_k \sigma_2 \langle g_k, p_k \rangle - \lambda_k \langle g_k, p_k \rangle \\
&= \lambda_k (\sigma_2 - 1) \langle g_k, p_k \rangle \\
&= \lambda_k (\sigma_2 - 1) \langle -(H_k)^{-1} p_k, p_k \rangle \\
&= \begin{cases} \lambda_k (\sigma_2 - 1) \langle -(H_k)^{-1} \tilde{p}_k, \tilde{p}_k \rangle & \text{if } \gamma > \gamma_{th} \\ \frac{\lambda_k}{\|\tilde{p}_k\|^2} (\sigma_2 - 1) \langle -(H_k)^{-1} \tilde{p}_k, \tilde{p}_k \rangle & \text{if } \gamma \leq \gamma_{th} \end{cases} > 0
\end{aligned}$$

where the first inequality comes from (W2), which implies

$$|\langle g_{k+1}, p_{k+1} \rangle| \leq \sigma_2 |\langle g_k, p_k \rangle| \quad (16)$$

and from the fact that  $\langle g_k, p_k \rangle = \langle -(H_k)^{-1} p_k, p_k \rangle < 0$  by assumption on  $H_k$  and the fact that the inverse matrix of a positive definite matrix is also positive definite; while the final strict positivity comes from the assumption  $\sigma_2 < 1$ .  $\square$

**Remark 1.** Note that Proposition 3 is not enough to prove a theoretical convergence result for the NNC-BFGS algorithm. Following [34], for that one would need to show there exists a positive constant  $\beta > 0$  such that the following relation holds

$$\|p_k\|_{H_k^{-1}}^2 \geq \beta \|r_k\|^2$$

for infinitely many  $k$ . In non-convex scenarios and for standard BFGS algorithms (without any cautious modifications of the form described in Section 5.1), this quantitative bound is well-known to be hard to prove. For the proposed NNC-BFGS update, the proof of this property and/or the use of a more cautious modification of the inner update is a matter of future research.

### 5.2.1. Numerical comparisons

We report an example of how the NNC-BFGS algorithm allows to reduce the computational complexity required to process all individual image patches, due, in particular to its dramatic improvement in reducing the number of iterations for challenging (i.e. associated to non-convex problems ( $P_y$ )) patches in the image. For our tests, we considered a set of 16 patches extracted from the OCT1 test image, as displayed in the left panel of Fig. 4. Each patch is assigned to a numbered label for referencing. These samples are indeed representative of the heterogeneity of image contents and, as a consequence, of the different regimes of non-convexity the method may have to deal with.

*Algorithmic parameters.* We compare the performance of the proposed NNC-BFGS with non-convexity threshold in (14) fixed as  $\gamma_{th} = 0.1$ , BFGS and CBFGS<sup>1</sup> with  $\varepsilon = 5$  in order to obtain a number of cautious updates that is consistent with the results in [25]. All methods include a line-search defined in terms of the strong Wolfe Conditions (W2), with a stopping criterion defined in terms of minimum step-size  $\lambda_{min} = 0.05$ .

Fig. 4 (right panel) proposes a visual insight of the sparsity of the estimated coefficient vectors  $a$  obtained by applying NNC-BFGS via an illustration of their histograms. As expected, all the histograms are rather centered around zero, and are spread to a higher or lesser extend depending on the particular image content considered (i.e. depending on the local estimated  $\gamma_y$ ). Fig. 5 shows the plots of the norm of the gradient  $\|\nabla f(a)\|_2$  in terms of the number of iterations for all the aforementioned methods when applied to the 16 test patches in Fig. 4. Note, in particular, that the last row shows the convergence behaviour of the algorithm when applied in patches characterised by a highly non-convex regime and is illustrative of the importance of carefully considering with this aspect. For all the remaining patches, the behaviour of NNC-BFGS corresponds to one of classic BFGS. Table 2 reports the number of iteration (nit) and the computational time required

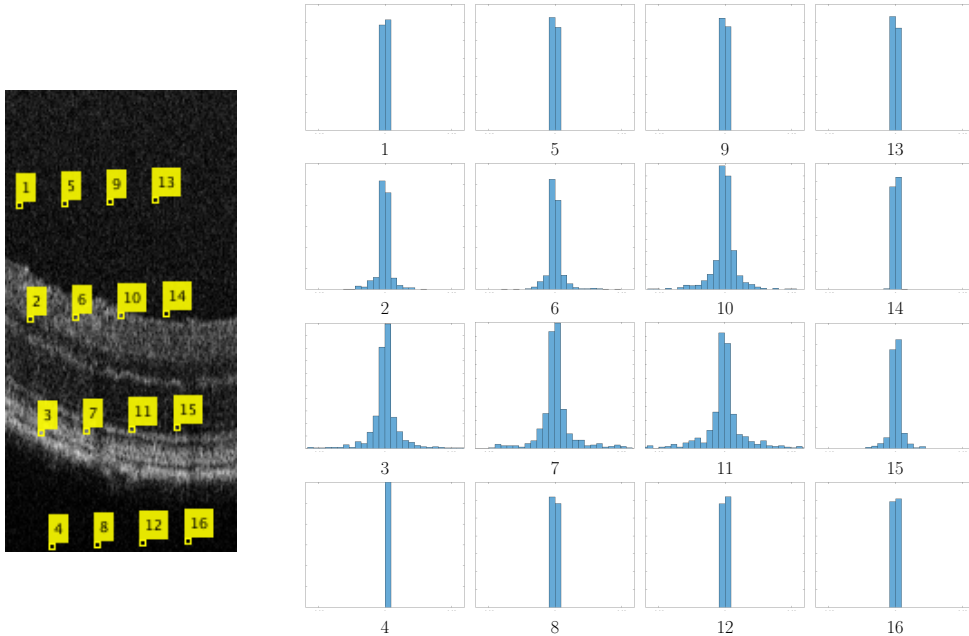


Figure 4: OCT1 image with 16 numbered representative patches (left); histograms of the corresponding estimated coefficient vectors  $a$  with horizontal axis spanning in the range  $[-0.4, 0.4]$  (right).

by all methods. In addition, for CBFGS it shows the number of cautious updates (cau) and for NNC-BFGS the value of  $\gamma$  computed for the patch under analysis. The NNC-BFGS outperforms both BFGS and CBFGS in the critical non-convex cases, identified by small  $\gamma_y$  as well as nonzeros cau values.

## 6. Numerical Tests

In Section 5.2.1 we presented results that validate the choice of the proposed space variant regularisation approach on a local level: for each subproblem  $(P_y)$  we managed to show that we can exploit the local information in order to obtain high quality results and design efficient optimisation strategies. In the current section we illustrate results that validate the overall SR reconstruction procedure sketched in Algorithm 1. Specifically, we performed experiments on real data, in the form of OCT images of murine retinas acquired by our collaborators within the AIR Lab at the University of Bristol. The images were acquired using a Micron IV system for imaging rodent eyes (Phoenix technologies, CA) and previously employed in [37, 5].

To build the dictionary  $D \in \mathbb{R}^{n_p \times n_d}$ , we used a sample of 60 noise-free HR OCT images as a training set for the SparseDT approach detailed in [32], under the assumption that the underlying data distribution is a Cauchy distribution. The choice of the dimensions of the dictionary -  $n_d = 600$  and  $n_p = 256$  - was driven by the results of our previous work [37], where several choices were performed and this particular combination led to the most promising results. We highlight that, provided a certain sought resolution for the HR image solution, our method is able to deal with different sampling factors  $q$  without the need to compute a new dictionary for each of them.

For all the experiments, we terminate the iterations of the NNC-BFGS algorithm in the inner loop of Algorithm 1 as soon as either of the two following conditions is satisfied

$$k > 5000, \quad \|\nabla f(x_k)\|_2 < 10^{-3}. \quad (17)$$

We will present three experimental results on the reconstruction of OCT images.

First, we validate the proposed super-resolution model for two different scaling factors  $q \in \{2, 4\}$  of the LR data. In the second example we will show a comparison with reconstructions obtained with other competing nonsmooth penalties and with/without the use of space-variant regularisation. Finally, in the third example, we will show how our reconstructions is well-suited

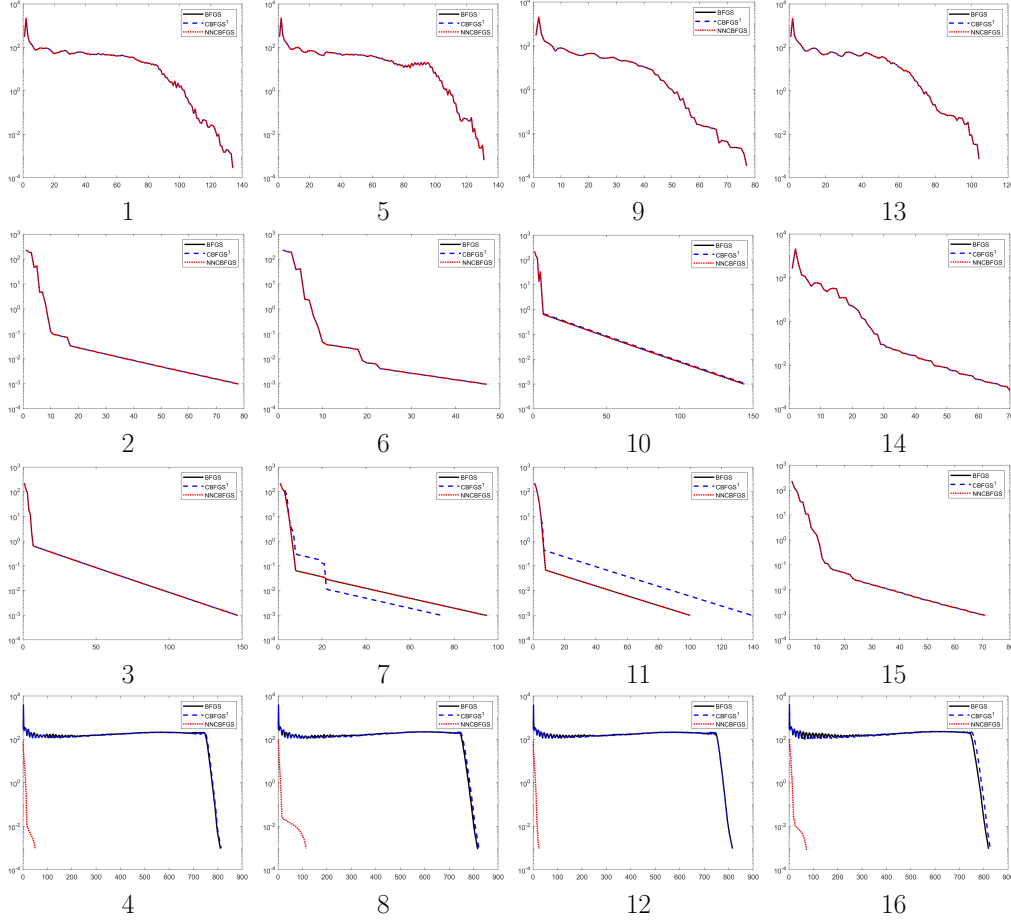


Figure 5: Gradient norm for selected patches illustrated in Fig.4.

to improve the quality of post-processing tasks such as  $k$ -means-based segmentation of the regions within OCT images.

### 6.1. OCT-SR reconstruction for different upscale factors $q$

In our first example, we employ Algorithm 1 with the NNC-BFGS optimisation solver to reconstruct an HR image starting from noisy and LR measurements synthetically generated by a downsampling of factor  $q = 2$  and  $q = 4$ .

Results are shown in Fig. 6, for the OCT images OCT3 (first row) and OCT4 (second row) and reported for the case  $q = 2$  (left panel) and  $q = 4$  (right panel). In each panel, column-wise, the figures show the  $\gamma$ -map estimated on the LR corresponding OCT image, the LR data and the recovered HR data, respectively.

As expected, the higher the dimension of the LR image (see case  $q = 2$ , where  $Y \in \mathbb{R}^{512 \times 256}$ ), the more detailed the  $\gamma$ -map will be, and, therefore, the more accurate the reconstructed HR OCT image ( $X^* \in \mathbb{R}^{1024 \times 512}$ ).

From a visual inspection of the  $\gamma$ -maps in Fig. 6 we can highlight how differently the structures in the vitreous are numerically treated. For  $q = 2$  many fragments of tissue suspended in the vitreous are identified. They are characterised by a high  $\gamma$  value, meaning that less sparsification is needed in those areas in order to obtain a richer reconstruction. On the contrary, for  $q = 4$ , only the main fragments are detected in the background with LR data. Nevertheless, this does not have an excessive impact on the quality of the reconstructed HR OCT image. Complementarily, low  $\gamma$  values in the bottom part of the  $\gamma$ -maps, force the non-convexity of the penalty in  $(P_y)$ , thus enforcing strong sparsity of solution, which leads to a predominantly zero reconstruction.

Patch	BFGS		CBFGS			NNC-BFGS		
	nit	time	nit	time	cau	nit	time	$\gamma$
1	134	7.86	134	7.75	0	134	7.54	0.14
2	78	4.86	78	4.66	0	78	4.69	0.59
3	147	9.12	147	9.10	2	147	8.85	0.92
4	814	49.45	814	50.05	2	<b>50</b>	<b>6.41</b>	0.09
5	77	4.36	77	4.34	0	77	4.27	0.17
6	144	8.69	146	8.53	1	144	8.86	0.86
7	100	5.91	139	8.61	3	100	6.05	0.95
8	815	47.94	816	49.86	3	<b>25</b>	<b>2.58</b>	0.09
9	104	5.89	104	6.14	0	104	5.97	0.15
10	70	4.1	70	3.98	0	70	3.95	0.23
11	71	4.09	71	3.88	0	71	3.96	0.45
12	819	49.5	827	49.2	12	<b>72</b>	<b>9.5</b>	0.08
13	81	4.52	81	4.53	0	81	4.6	0.19
14	116	6.7	116	6.55	0	116	6.94	0.15
15	61	3.45	61	3.4	0	61	3.36	0.26
16	820	50.27	828	49.74	10	<b>54</b>	<b>6.62</b>	0.08

Table 2: Computational efficiency of BFGS, CBFGS and NNC-BFGS in terms of number of iterations (nit) and computational time (in seconds) for the 16 patches in Fig. 4; for CBFGS we also report the number of cautious updates throughout the iterations, while for NNC-BFGS we report the value of the estimated  $\gamma$  for the current patch.

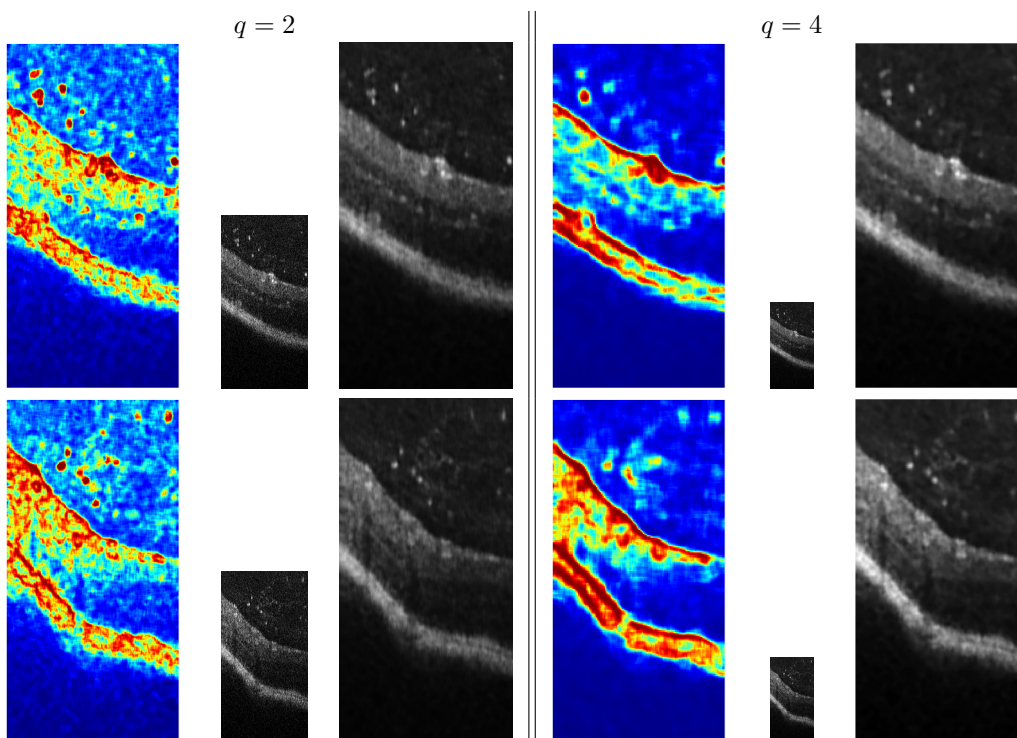


Figure 6: Reconstruction for OCT3 (first row) and OCT4 (second row) with scaling factor  $q = 2$  (left panel) and  $q = 4$  (right panel). Left panel, from left to right:  $\gamma$ -map ( $512 \times 256$ ), LR data ( $512 \times 256$ ), HR reconstruction ( $1024 \times 512$ ). Right panel, from left to right:  $\gamma$ -map ( $256 \times 128$ ), LR data ( $256 \times 128$ ), HR reconstruction ( $1024 \times 512$ ).

### 6.2. Comparison between different regularisation models

We now investigate how the choice of the penalty function in the inner minimisation problem ( $P_y$ ) affects the overall reconstruction. Namely, we consider as a non-smooth convex penalty  $f_2(\cdot) = \lambda \|\cdot\|_1$ ,  $\lambda > 0$ , and as a non-smooth non-convex penalty, the MCP function defined as

$$\psi(t; \alpha) := \begin{cases} -\frac{1}{2\alpha}t^2 + \sqrt{\frac{2}{\alpha}}t & \text{if } |t| < \sqrt{2\alpha}, \\ 1 & \text{if } |t| \geq \sqrt{2\alpha}, \end{cases} \quad (\text{MCP})$$

where the parameter  $\alpha > 0$  modulates the concavity of the regulariser [44]. We thus compare the variational problem ( $P_y$ ) where the Cauchy penalty is used - denoted in the following as ( $P_y - Ca$ ) - with the analogous penalty-dependent problems:

$$a_y^* \in \operatorname{argmin}_{a \in \mathbb{R}^{n_d}} \left\{ f(a) := \frac{1}{2} \|y - \mathcal{S}_q \mathcal{KD}(a)\|_2^2 + \lambda \sum_{i=1}^{n_d} |a_i| \right\}, \quad (P_y - \ell_1)$$

$$a_y^* \in \operatorname{argmin}_{a \in \mathbb{R}^{n_d}} \left\{ f(a) := \frac{1}{2} \|y - \mathcal{S}_q \mathcal{KD}(a)\|_2^2 + \lambda \sum_{i=1}^{n_d} \psi(a_i; \alpha) \right\}. \quad (P_y - \text{MCP})$$

The regularisation parameter  $\lambda$  is globally fixed throughout the image. In this manner the effect of the penalty is homogeneous throughout the image according to the given  $\lambda$  value, thus yielding either excessive or insufficient regularisation in the different patches, contrarily to what is advisable to deal with the high inhomogeneity of the considered OCT images. The tuning of the  $\lambda$  value can be entrusted to standard selection strategies relying, for instance, on the discrepancy-principle, Generalised Cross Validation or more sophisticated statistical approaches. However, those strategies often rely on previous knowledge of either the noise level in the data and/or computationally expensive matrix computations. On the other hand, in ( $P_y - Ca$ ) the value  $\gamma_y$  is estimated directly from the data and allows for the design of an adaptive-regularisation tuning for each subproblem ( $P_y$ ).

Note that both non-smooth optimisation problems ( $P_y - \ell_1$ ) and ( $P_y - \text{MC}$ ) can be efficiently solved by means of proximal gradient strategy, as explained in [37].

The reconstructions obtained by means of these different regularisers are compared on OCT1 image (Fig. 7, first row) and OCT2 image (Fig. 7, second row) for an upsampling factor  $q = 4$ . The figure is divided into three panels: the left panel shows the HR reconstruction obtained with the proposed Space Variant (SV) approach. The other HR OCT-image reconstructions illustrated in the centered and right panels are obtained by running Algorithm 1 with optimisation problems ( $P_y - \ell_1$ ) and ( $P_y - \text{MC}$ ), respectively. For both  $\ell_1$  and MCP regularisers, we report the reconstructed OCT images obtained using a global high and low  $\lambda$  value. For high values of  $\lambda$  a good removal of the background noise is obtained, but, on the other hand, several details in the central layered part of the OCT image are smoothed out. On the other hand, for a small value of  $\lambda$  an accurate reconstruction of the textured layers is obtained, which, however, is unable to remove the background noise at the same time.

### 6.3. OCT-image segmentation

The poor spatial resolution and the noise observed in OCT images often limits the possibility of an accurate subsequent image analysis, often based on a time-consuming manual layer segmentation [41, 6]. Here, we illustrate how the use of the proposed reconstruction model may improve the quality in OCT image segmentation. To make our point, we consider a simple  $k$ -means-based partitioning approach, where the value  $k = 4$  was selected so as to represent different regions in the image.

Recalling Section 4, one might think of obtaining a first partitioning result directly from the analysis of the  $\gamma$ -map range, since in our estimation procedure the values  $\gamma_{i,j}$  are intended to describe the inhomogeneity within the data. We illustrate an example of this preliminary test in Fig. 8, where on the left we report the  $\gamma_{i,j}$  values estimated by means of the the procedure detailed in Section 4, while on the right we illustrate its  $k$ -means segmentation, with  $k = 4$ . As it can be noticed, such segmentation properly separates the main DS regions in the image and spots some



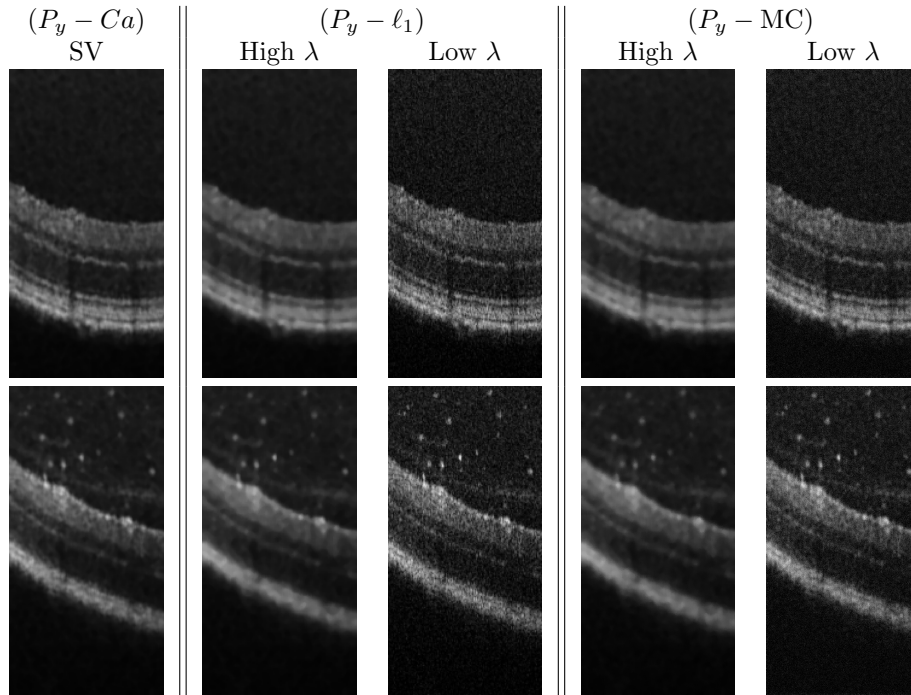


Figure 7: Effects of different regularisers in  $(P_y)$  on the OCT image reconstruction.

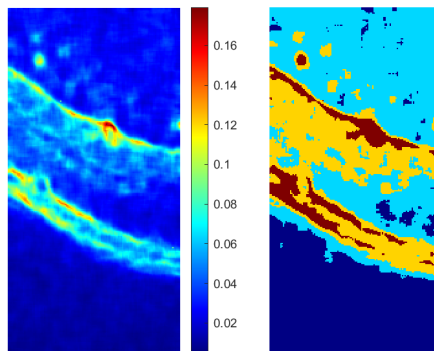


Figure 8: Segmentation:  $\gamma_{i,j}$  values of a LR OCT image (left) and their  $k$ -means partitioning with  $k = 4$  (right)

of the fragments of tissue suspended in the upper part. However, it is worth recalling that such partitioning is obtained by LR OCT data hence it may lack of more detailed HR contents.

Better results can be obtained by performing image partitioning on the HR solutions obtained by the proposed method, yielding in particular higher precision in detecting the layered region within the retina and the portions of tissue suspended in the vitreous. We illustrate this in Fig. 9 where the  $k$ -means segmentation has been applied to OCT2, OCT3 and OCT4 images. Fig. 9 is divided into two panels: the left one is dedicated to the LR data with the associated segmentation results, while the right one shows the HR reconstructions obtained by Algorithm 1 and their segmentations into  $k = 4$  regions. Segmentation on HR OCT images definitely improves the segmentation obtained with simple LR images. The application of the adaptive SR preliminary process allowed us to achieve remarkable results in separating the layers of the retinal structure, with simultaneous background noise removal.

## 7. Conclusions

We proposed an adaptive, smooth, non-convex and sparsity-promoting variational approach for the reconstruction of high-resolution OCT data from low-resolution measurements using a sparse

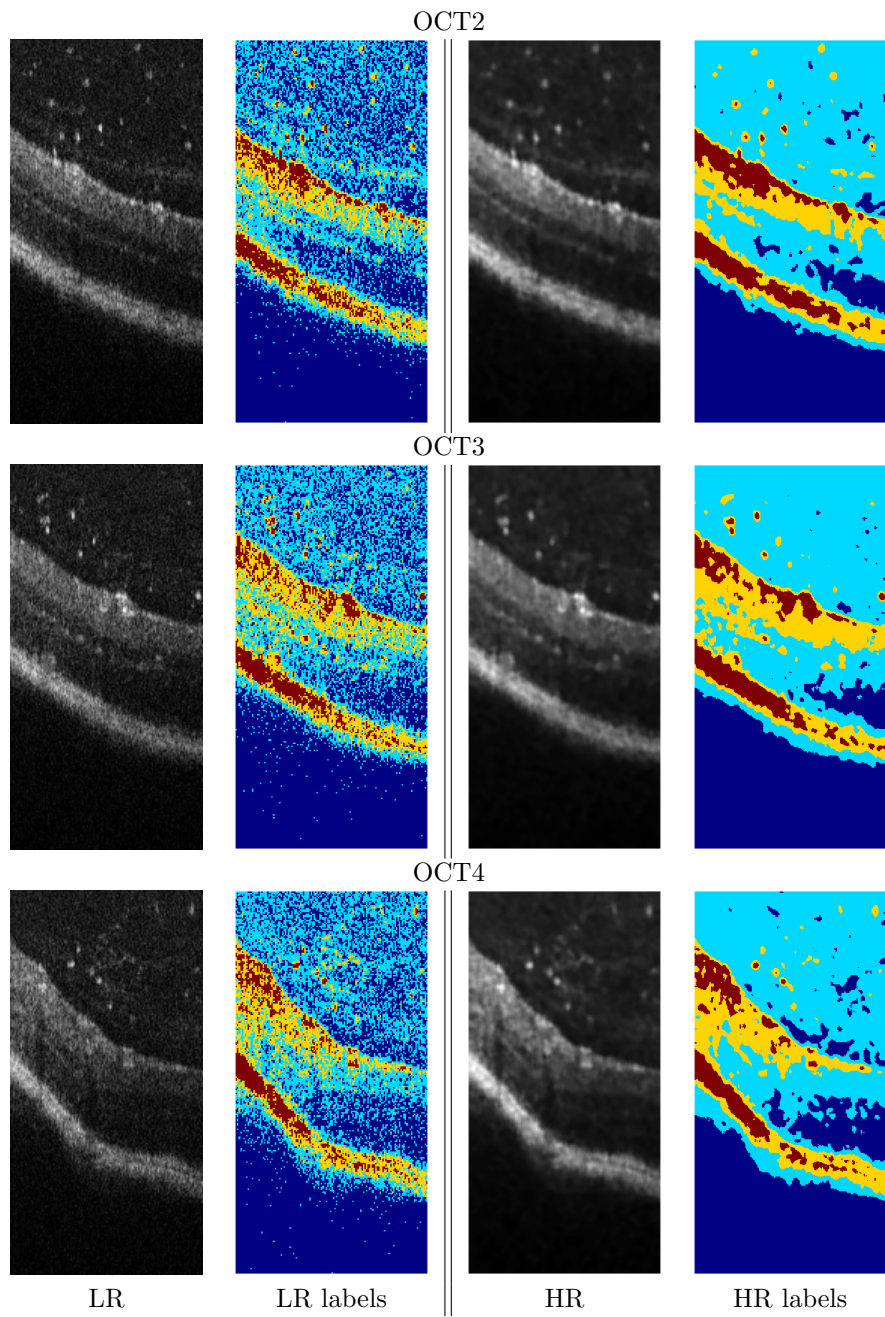


Figure 9:  $k$ -means segmentation with  $k = 4$  of OCT2, OCT3 and OCT4

representation framework. Motivated by statistical considerations, a Cauchy-based penalty was used to promote sparsity. Differently from more standard regularisation models classically used as relaxation of the intractable  $\ell_0$  pseudo-norm, such penalty depends on a scale parameter whose size determines not only the regularisation strength, but also its shape ranging from convex to non-convex. A patch-adaptive parameter estimation strategy based on Maximum Likelihood is proposed which allows to take into account the heterogeneous content observed in real OCT data. The strategy is shown to produce an adaptive map of parameters which self-tune the regularisation depending on the type of local information observed. In order to exploit the overall smoothness of the resulting variational model obtained when combined the proposed Cauchy penalty with an  $\ell_2$  data term, we considered a variant of the well-known BFGS algorithm endowed with a cautious update of the descent direction and provided several numerical tests showing improved computational efficiency. We validated the overall SR framework on different real OCT images. We showed the advantages of having an adaptive regularisation model with respect to a fixed one, and compared the results with the ones obtained with popular non-smooth convex and non-convex regularisers frequently employed to enforce sparsity. Finally, we showed how the reconstructed data can be successfully employed to extract hidden information for image segmentation purposes via a standard  $k$ -means algorithm. Image classification is indeed a crucial task in the analysis of OCT data in particular and we expect that the proposed pipeline could indeed be beneficial for further studies combined with more sophisticated image segmentation strategies.

## Acknowledgements

The authors are grateful to Dr Lindsay Nicholson (University of Bristol, UK) and his colleagues within the Autoimmune Inflammation Research Lab, for providing the murine OCT datasets used to validate this study. The work of Achim was supported in part by a fellowship from the Leverhulme Trust (INFHER). Calatroni acknowledges the support received by the Marie Skłodowska-Curie project H2020 RISE NoMADS, GA 777826 and by the grant DEP “Attrattività du territoire” by UCA IDEX JEDI. Scrivanti acknowledges the financial support from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861137. Research by Morigi was supported in part by the National Group for Scientific Computation (GNCS-INDAM), Research Projects 2021.

## References

- [1] A. Achim, A. Bezerianos, and P. Tsakalides. Novel Bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Trans. Med. Imag.*, 20:772–783, Aug. 2001.
- [2] A. Achim and E. Kuruoglu. Image denoising using bivariate  $\alpha$ -stable distributions in the complex wavelet domain. *IEEE Signal Processing Letters*, 12(1):17–20, 2005.
- [3] A. Achim, P. Tsakalides, and A. Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. and Remote Sensing*, 41:1773–1784, Aug. 2003.
- [4] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.
- [5] L. J. Bradley, A. Ward, M. C. Y. Hsue, J. Liu, D. A. Copland, A. D. Dick, and L. B. Nicholson. Quantitative assessment of experimental ocular inflammatory disease. *Frontiers in Immunology*, 12:2308, 2021.
- [6] X. Cai, R. Chan, S. Morigi, and F. Sgallari. Vessel segmentation in medical imaging using a tight-frame-based algorithm. *SIAM Journal on Imaging Sciences*, 6(1):464–486, 2013.
- [7] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari. A flexible space-variant anisotropic regularization for image restoration with automated parameter selection. *SIAM J. Imaging Sci.*, 12:1001–1037, 2019.

- [8] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari. Adaptive parameter selection for weighted-TV image reconstruction problems. In *J.Phys.: Conf. Series, NCMIP 2019*, volume 1476, pages 541–547, 2020.
- [9] D. Calvetti, M. Pragliola, E. Somersalo, and A. Strang. Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors. *Inverse Problems*, 36(2), jan 2020.
- [10] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52(2):489–509, Feb. 2006.
- [11] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001.
- [12] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, 2011.
- [13] V. Das, S. Dandapat, and P. K. Bora. A diagnostic information based framework for super-resolution and quality assessment of retinal oct images. *Computerized Medical Imaging and Graphics*, 94:101997, 2021.
- [14] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.
- [15] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE transactions on medical imaging.*, 32(11):2034–2049, 2013-11.
- [16] Y. Huang, Z. Lu, Z. Shao, M. Ran, J. Zhou, L. Fang, and Y. Zhang. Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network. *Opt. Express*, 27(9):12289–12307, Apr 2019.
- [17] M. Huska, D. Lazzaro, S. Morigi, A. Samorè, and G. Scrivanti. Spatially-adaptive variational reconstructions for linear inverse electrical impedance tomography. *Journal of Scientific Computing*, 84(3):46, 2020.
- [18] O. Karakuş and A. Achim. On solving sar imaging inverse problems using nonconvex regularization with a cauchy-based penalty. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5828–5840, 2021.
- [19] O. Karakuş, N. Anantrasirichai, A. Aguersif, S. Silva, A. Basarab, and A. Achim. Detection of line artifacts in lung ultrasound images of covid-19 patients via nonconvex regularization. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(11):2218–2229, 2020.
- [20] O. Karakuş, P. Mayo, and A. Achim. Convergence guarantees for non-convex optimisation with cauchy-based penalties. *IEEE Transactions on Signal Processing*, 68:6159–6170, 2020.
- [21] A. Lanza, S. Morigi, M. Pragliola, and F. Sgallari. Space-variant generalised gaussian regularisation for image restoration. *Comput. Meth. Biomech. Biomed. Eng.: Imaging Vis.*, 13:490–503, 2018.
- [22] A. Lanza, S. Morigi, M. Pragliola, and F. Sgallari. Space-variant TV regularization for image restoration. In J. M. R. Tavares and R. Natal Jorge, editors, *VipIMAGE 2017*, pages 160–169, Cham, 2018. Springer Intern.Pub.
- [23] A. Lanza, S. Morigi, I. W. Selesnick, and F. Sgallari. Sparsity-inducing nonconvex non-separable regularization for convex image processing. *SIAM Journal on Imaging Sciences*, 12(2):1099–1134, 2019.

- [24] A. Lanza, S. Morigi, and F. Sgallari. Convex image denoising via non-convex regularization with parameter selection. *Journal of Mathematical Imaging and Vision*, 56:195–220, 2016.
- [25] D.-H. Li and M. Fukushima. On the global convergence of the bfgs method for nonconvex unconstrained optimization problems. *SIAM Journal on Optimization*, 11(4):1054–1064, 2001.
- [26] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:674–692, July 1989.
- [27] P. McCullagh and N. G. Polson. Statistical sparsity. *Biometrika*, 105(4):797–814, 10 2018.
- [28] O. V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 53(1):64–78, 2006.
- [29] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.
- [30] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley and Sons, New York, 1995.
- [31] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.
- [32] P. Pad, F. Salehi, E. Celis, P. Thiran, and M. Unser. Dictionary learning based on sparse distribution tomography. volume 70 of *Proceedings of Machine Learning Research*, pages 2731–2740. PMLR, 2017.
- [33] M. Pereyra and S. McLaughlin. Fast unsupervised bayesian image segmentation with adaptive spatial regularisation. *IEEE Trans. Image Proc.*, 26:2577–2587, 2017.
- [34] M. J. D. Powell. On the Convergence of the Variable Metric Algorithm. *IMA Journal of Applied Mathematics*, 7(1):21–36, 02 1971.
- [35] P. Puvanathasan and K. Bizheva. Speckle noise reduction algorithm for optical coherence tomography based on interval type II fuzzy set. *Opt. Express*, 15(24):15747–15758, Nov 2007.
- [36] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.
- [37] G. Scriveranti, L. Calatroni, S. Morigi, L. Nicholson, and A. Achim. Non-convex super-resolution of oct images via sparse representation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 621–624, 2021.
- [38] E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for  $\ell_2$ - $\ell_0$  minimization. *SIAM Journal on Optimization*, 27(3):2034–2060, 2017.
- [39] R. Svetlozar T. In S. T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages ix–xi. North-Holland, Amsterdam, 2003.
- [40] M. Unser and P. D. Tafti. *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2014.
- [41] Q. Wang, R. Zheng, and A. Achim. Super-resolution in optical coherence tomography. In *40th Annual IEEE EMBS Conference*, pages 1–4, 2018.
- [42] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Process.*, 19(11):2861–2873, 2010.
- [43] D. V. Zermeno, P. Mayo, L. Nicholson, and A. Achim. Super-resolution oct using sparse representations and heavy-tailed models. In *41st Annual IEEE EMBS Conference*, pages 5585–5588, 2019.
- [44] C.-H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.*, 38(2):894–942, 04 2010.

- [45] A. M. Zysk, F. T. Nguyen, A. L. Oldenburg, D. L. Marks, and S. A. B. M.D. Optical coherence tomography: a review of clinical development from bench to bedside. *Journal of Biomedical Optics*, 12(5):1 – 21, 2007.