

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

Alin Achim^{*#} Luca Calatroni^{*} Serena Morigi[†] Gabriele Scrivanti^{*§}

[†] *Department of Mathematics, University of Bologna, P.zza Porta S. Donato 5, Italy.*

^{*} *Laboratoire I3S, CNRS, UCA, INRIA, 2000 Route des Lucioles, 06903 Sophia-Antipolis, France*

[#] *Visual Information Laboratory, University of Bristol, Bristol, UK.*

[§] *Université Paris-Saclay, Inria, CentraleSupélec, Centre de Vision Numérique, Gif-sur-Yvette, France.*

Abstract

We propose a smooth, non-convex and content-adaptive regularisation model for single-image super-resolution of 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 a pre-learned dictionary tailored to model α -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 separable Cauchy-type penalty. To favour adaptivity to image contents, we propose a space-variant modelling by which the local degree of non-convexity given by the local Cauchy shape parameter is estimated via maximum likelihood. For the solution of the reconstruction problem, 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 of the functional. Our numerical results show that the combination of a space-variant modelling with a tailored optimisation strategy improves reconstruction results and allows for an effective segmentation with standard approaches.

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., [31, 4, 13, 48]) but also interpreted within a statistical setting and in terms of heavy-tailed probability density functions [44]. As far as the use of sparse representation models for the specific problem of super-resolution is concerned, in their seminal work [51], 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 [40, 12, 25, 26, 10, 11]. The notion of sparsity can be then described in terms of the minimisation of the non-convex and non-continuous ℓ_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 [36], so that approximated approaches have to be considered in order to solve the problem. A popular choice consists in replacing the ℓ_0 pseudo-norm with the convex ℓ_1 norm [14, 18], 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 ℓ_p “norms” ($p < 1$), Smoothly Clipped Absolute Deviation, Minimax Concave Penalty (MCP) and Continuous Exact relaxations, see [46] 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., [32] 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 [44, 37]. 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 generally empirically selected, the use of alpha-stable densities is rigorously justified by the generalised central limit theorem [44]. 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 [24, 22, 23]. The smoothness of such penalty is particularly interesting from an optimisation point of view as it allows in principle a handy numerical treatment via gradient-based solvers.

Following [24, 22, 23], 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., [54] 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 [52, 50]. 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 [16] 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. As far as previous approaches are concerned, in [19] the authors proposed a sparsity-based simultaneous denoising and interpolation of OCT images, while in [16] 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 [20]. 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.

Contributions. We propose a mathematical approach for the reconstruction of OCT data using the framework of sparse-representation [51] and a Cauchy-type regulariser to promote sparsity. Our proposal extends the one preliminarily studied in [45] with respect to:

- (i) the design of a new sparse-representation-based reconstruction approach with focus on a super-resolution framework;

- (ii) the use of a patch-adaptive Cauchy-type regularisation for the estimation of relevant weights and combined with an automatic parameter estimation strategy;
- (iii) the proposal of quasi-Newton optimisation scheme adapted to the local convex/non-convex behaviour and computing effective model solutions.

In details:

- (i) The proposed design of a reconstruction approach for OCT images is based on their sparse representation with respect to a fixed high-resolution (HR) dictionary performed on a small set of properly denoised HR images. Compared to classical approaches, see, e.g., [51], the main advantage of the proposed approach is that it does not require the learning of a low resolution (LR) dictionary, while allowing at the same time the reconstruction of LR data independently on their initial resolution. Since the reference dictionary is built on HR cleaned data, an intrinsic denoising effect within the reconstruction process is also promoted.
- (ii) The patch-wise heterogeneity of image contents is taken into account by the design of a patch-adaptive Cauchy regularisation term whose local degree of regularisation and non-convexity is adapted to the local image statistics. The proposed estimation strategy relies entirely on the use of the given LR data. As such, it can be performed offline making the proposed approach *parameter-free*. While sharing few analogies with the space-variant regularisation approaches (see, e.g., [21, 42] and references therein) the proposed strategy is significantly different from others used in space-variant noise analysis (see, e.g., [5, 17, 7]) where the estimation is done to better describe heterogeneous noise (rather than solution) statistics.
- (iii) The smoothness of the composite reconstruction model considered is exploited by considering a cautious modification of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. Differently from previous work [24, 45], where the Cauchy penalty is manipulated in terms of its proximal operator and within a first-order proximal gradient perspective, here we exploit the smoothness of such penalty and introduce an efficient Non-Convex Cautious BFGS (NNC-BFGS) algorithm tuning its update rules depending on the local degree of convexity estimated from the data as in (ii).

We stress that 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 technique proposed, e.g., in [51]. Note, however, that our approach differs from standard approaches from two main perspectives: firstly, it relies upon a single HR dictionary and not on a LR-HR pair; secondly, our reconstruction process favours an intrinsic denoising process which is, in general, not a feature of classical SR techniques.

Structure of the paper. 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 γ -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 [51] 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

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

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 [43]. 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 ℓ_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 ℓ_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 γ -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 [24, 45]. The patch-dependent parameter γ_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 [45], 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 κ) or estimated as described in [33].

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 [51]. To build the HR dictionary \mathcal{D} , we follow [39] where general α -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 Π_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 Π_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 Ω_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, K, S_q

Output: X^* % HR OCT image

Estimate the γ -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 Π_q

Compute median values of all overlapping pixels

Collect all the median values in X^*

The computational complexity of Algorithm 1 boils down to two bottlenecks: the offline and resolution-independent estimation of the γ -map (which involves the LR data only and it is thus performed before the reconstruction step) and the solution to problem (P_y) for each patch. In Section 4 we comment on the former step and include some insights on the computational times required to build such map for a test image. As far as the reconstruction algorithm complexity is concerned, in Section 5, we report the computational time required to solve the problem (P_y) for different patches y , extracted from a test image.

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) .

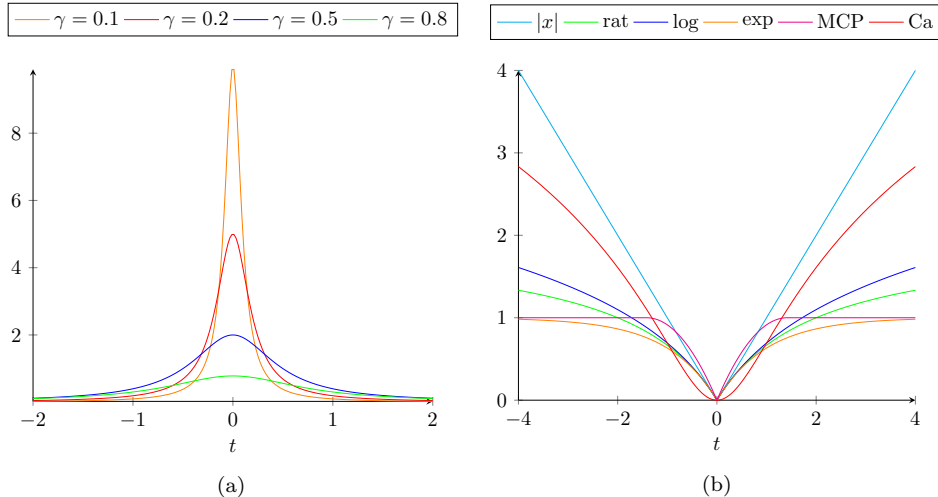


Figure 1: (a) Cauchy PDF for varying $\gamma > 0$. (b) Plots of different penalty functions relaxations of the ℓ_0 pseudo-norm.

From a statistical point of view, the Cauchy distribution belongs to the family of α -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 [47, 52, 2]. More precisely, a Cauchy distribution is an α -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:

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

As illustrated in Fig. 1a the value of γ controls the spread of the Cauchy distribution, playing thus the very same role as the Gaussian variance: the smaller γ , 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 ℓ_1 norm and other popular penalties typically used as alternatives to ℓ_0 pseudo-norm, compare e.g. with [28, 46]. Note that the function $\phi(\cdot; \gamma)$ in (2) is non-convex except for a small and limited interval around the origin, and increases unboundedly 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 $\mathcal{H}_f \in \mathbb{R}^{n_a \times n_a}$ of the function $f(a)$ in (P_y) is indefinite and reads as follows

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

where $\phi''(\cdot)$ is defined as in (3), and $S_q K D \in \mathbb{R}^{\frac{n_p}{q^2} \times n_a}$. 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 ℓ_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 [45]. 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 [27]. However, its lack of smoothness forces to the use of non-smooth optimisation methods such as proximal point algorithms [15, 21] to deal with it from a numerical point of view. Analogous considerations hold for ℓ_1 regularisation. On the other hand, according to Proposition 1, the function ϕ 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 γ guaranteeing that the associated functional in (P_y) is convex as done, e.g., in [27].*

4. Design of the space-variant γ -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 γ -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 γ which can be written as

$$\gamma_{i,j} \in \arg \max_{\gamma} \sum_{(\ell,\kappa) \in \mathcal{N}_{i,j}} \log p(\tilde{y}_{\ell,\kappa}; \gamma) \quad (5)$$

where p is the PDF in (1). The expression in (5) can be minimised using any unconstrained function optimisation method, such as the Nelder-Mead simplex direct search approach. In our implementation we have used Matlab's `fminsearch` function. As far as the computational cost of such procedure is concerned, the average CPU time to estimate the values $\gamma_{i,j}$ based on a 9×9 neighbourhood for each single pixel $(i, j) \in \Omega_l$ of a 256×128 image is 0.0027 seconds. The total cumulative time sums up to 90.2026 seconds.

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 γ -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 γ in the penalty ϕ 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

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

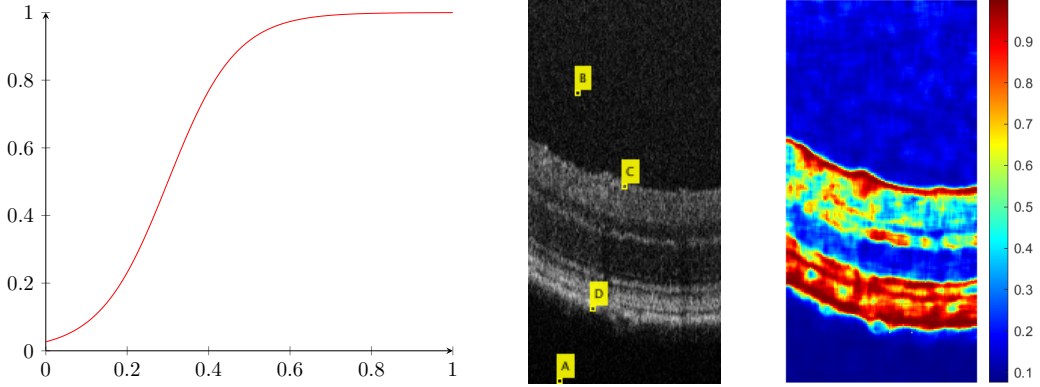


Figure 2: Rescaling sigmoid function (left); LR OCT image with four different labelled patches (centre); corresponding γ -map (right).

which is the strictly increasing function illustrated in Fig. 2(left). Finally, the value for γ_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 γ -map associated to the LR OCT image in the middle, with values $\gamma_{i,j} \in [0.1, 1]$ for all i, j .

To better highlight the link between the γ -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 γ_y is extracted from the corresponding γ -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 γ values, fixed for each patch, in the range $\gamma \in \{0.1, 0.2, 0.5, 0.9\}$, respectively. As the parameter γ 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 γ and space-variant γ_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 γ parameter, thus reducing the computational efforts without affecting (if not improving) the reconstruction quality.

# Patch \ γ	0.1	0.2	0.5	0.9	SV γ
A (background)	38.053	38.041	37.886	37.584	38.055 (0.08)
B (vitreous)	24.157	24.172	24.137	24.046	24.162 (0.14)
C (layers)	25.642	26.183	27.295	27.504	27.525 (0.86)
D (layers)	17.264	18.651	22.840	24.957	25.072 (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 γ_y where the estimated γ_y values are reported in brackets. Best PSNR values are boldfaced.

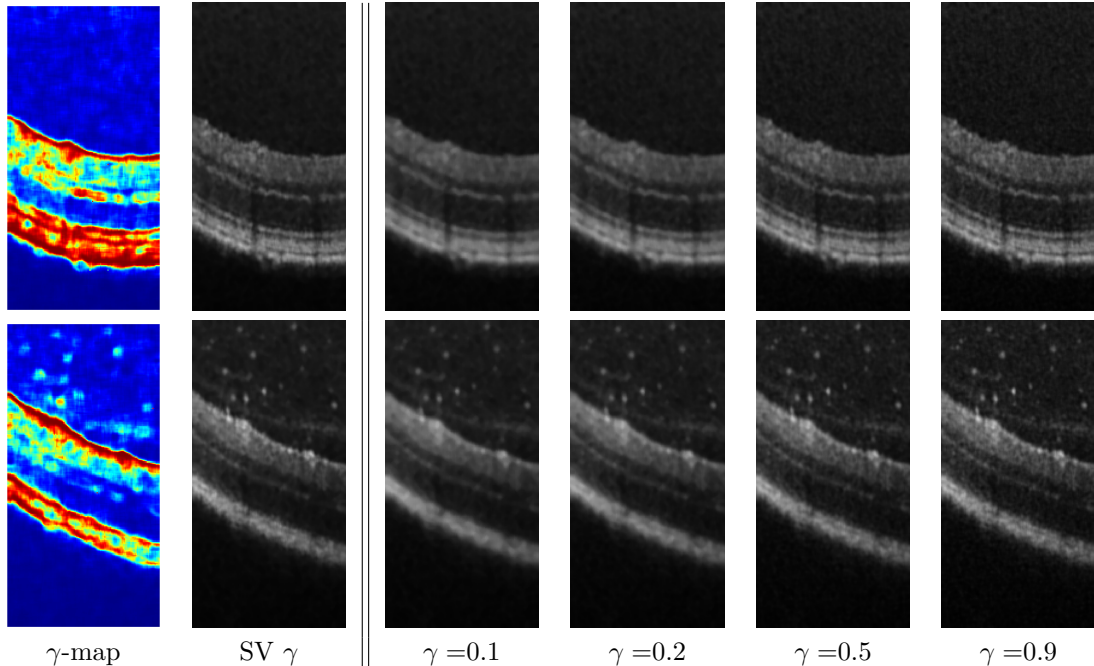


Figure 3: Comparisons between the proposed space-variant (SV) reconstruction approach (left) and the one with fixed values of γ (right) on the images OCT1 (first row) and OCT2 (second row), obtained with an upscaling factor $q = 4$.

5. Smooth, non-convex optimisation via cautious BFGS

From an optimisation viewpoint, the patch-dependent problem (P_y) is small-sized and smooth, although possibly non-convex. The gradient $\nabla f(\cdot)$ of the cost function f in (P_y) is Lipschitz continuous with constant $L > 0$ which can be estimated by triangle inequality as $L \leq L_{f_1} + L_{f_2}$ where L_{f_1} and L_{f_2} are, respectively, the Lipschitz constants of the quadratic 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, which, compared to first-order (proximal) gradient and ADMM algorithms can achieve faster (super-linear) convergence in convex regimes, thus improving overall efficiency. Furthermore, since the dimension of the patch sub-problem is low, the computational costs required to compute an approximation of the Hessian matrix are reduced.

Recalling the range of convexity of the Cauchy penalty as in Proposition 1, we deduce that an iterative procedure for solving the composite problem (P_y) might shift along the iterations from a convex ($\|a_y^k\|_\infty \leq \gamma_y$) to a non-convex ($\|a_y^k\|_\infty > \gamma_y$) regime, depending on the values of the vector a^k of the estimated coefficients at the current iteration k . For patches characterised by small values of γ_y , the range of convexity is very narrow and a careful strategy is required to keep the values within the desired interval to favour the algorithm to benefit from super-linear convergence therein.

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 [38], its global behaviour in a non-convex setting is still an open question.

In [30] 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 ε and α in (12) clearly affects the practical application of the update rules. In [30] the authors suggest to choose a fixed ε and to either vary α according to the following criterium

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

or to set $\alpha = 1$ (CBFGS²).

5.2. Normalised NonConvex BFGS

By Proposition 1, problem (P_y) is convex in the range $[-\gamma, \gamma]$. Outside of this range, the standard BFGS method is not guaranteed to converge. Taking advantage of the information encoded in the γ -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. Instead of the strategy (12)-(CBFGS¹), 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) thus simply consists in normalising the descent direction so as to allow the inexact linesearch to look for an overall appropriate step-size λ_k in (10). For values $\gamma \leq \gamma_{th}$ the Cauchy penalty $\phi(a; \gamma)$ is likely to have a narrow range of convexity with a second derivative with great magnitude affecting the nonlinear part of the Hessian matrix as in (4). The scaling on p_k thus 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 λ_k satisfies 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 [41], for that one would need to show that 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.

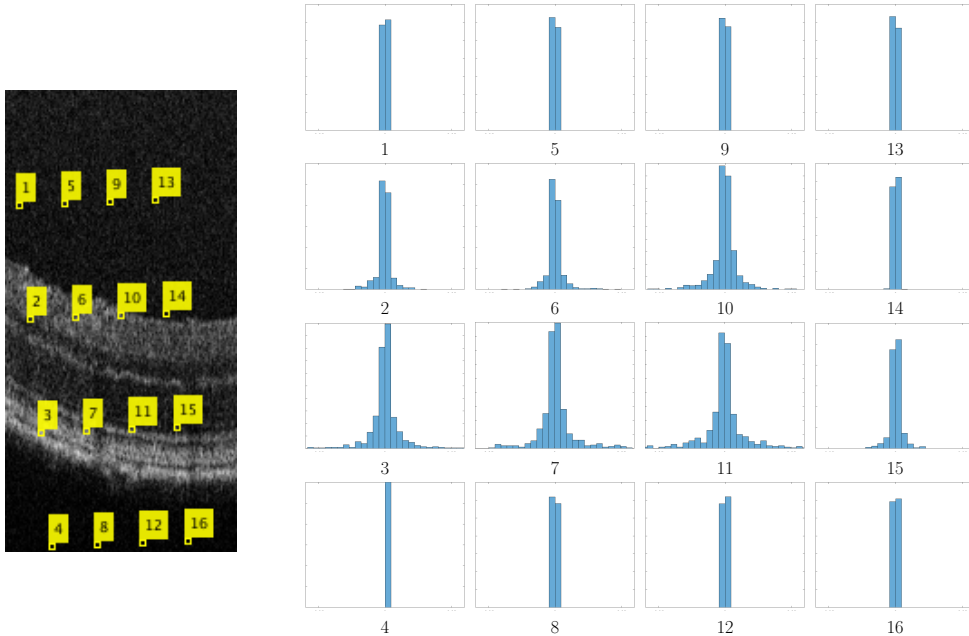


Figure 4: OCT1 image with 16 numbered representative patches (left); histograms of the corresponding estimated coefficient vectors a with horizontal axis spanning in the range $[-0.4, 0.4]$ (right).

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 CBFGS¹ with $\varepsilon = 5$ in order to obtain a number of cautious updates that is consistent with the results in [30]. 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 γ_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 narrow convexity range and is illustrative of the importance of carefully considering with this aspect. For all the remaining patches, the behaviour of NNC-BFGS corresponds to the one of classic BFGS. Table 2 reports the number of iteration (nit) and the computational time required by all methods. In addition, for CBFGS it shows the number of cautious updates (*cau*) and for NNC-BFGS the value of γ computed for the patch under analysis. The NNC-BFGS outperforms both BFGS and CBFGS in the critical cases identified by small γ_y as well as nonzero *cau* values. Preliminary comparisons with plain gradient descent (GD) optimisation showed that convergence for GD up to the desired tolerance is not always guaranteed within the fixed maximum number of iterations (1000). For this reason, we decided not to add such poorly informative convergence plots. For the same patches illustrated in Fig. 4 we also report in Fig. 6 (in blue) the maximum absolute coefficient value $\|a_y^k\|_\infty$ throughout the iterations of NNC-BFGS $k = 1, 2, \dots$. The red dashed line represents the threshold γ_{th} defined in (14), while the red solid line represents the mean value of γ_y defined in equation (6) for the corresponding patch y . As noted at the beginning of this section, the condition $\|a_y^k\|_\infty \leq \gamma_y$ is sufficient to ensure convexity of the objective function for each patch. This is something that thanks to the proposed NNC-BFGS we are able to control and which shows clear advantages, e.g., in the lower patches of the image, as shown in the last row of plots in Fig. 6, where the convexity condition is always satisfied.

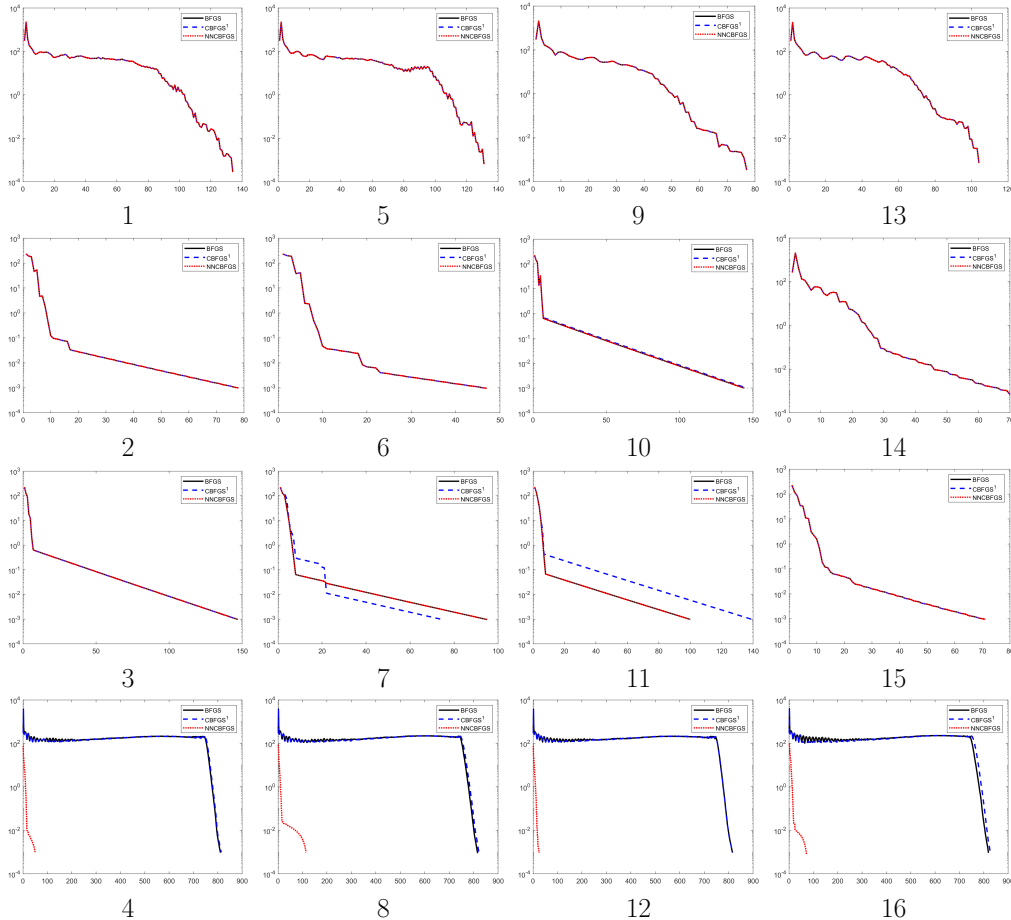


Figure 5: Gradient norm for selected patches illustrated in Fig.4. Super-linear convergence is achieved by the proposed algorithm for patches 4, 8, 12 and 16 due to the local convexity enforcement.

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 [45, 8].

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 [39], 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 [45], where several choices were considered 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

$$k > 1000, \quad \|\nabla f(x_k)\|_2 < 10^{-3}. \quad (17)$$

We will present four 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

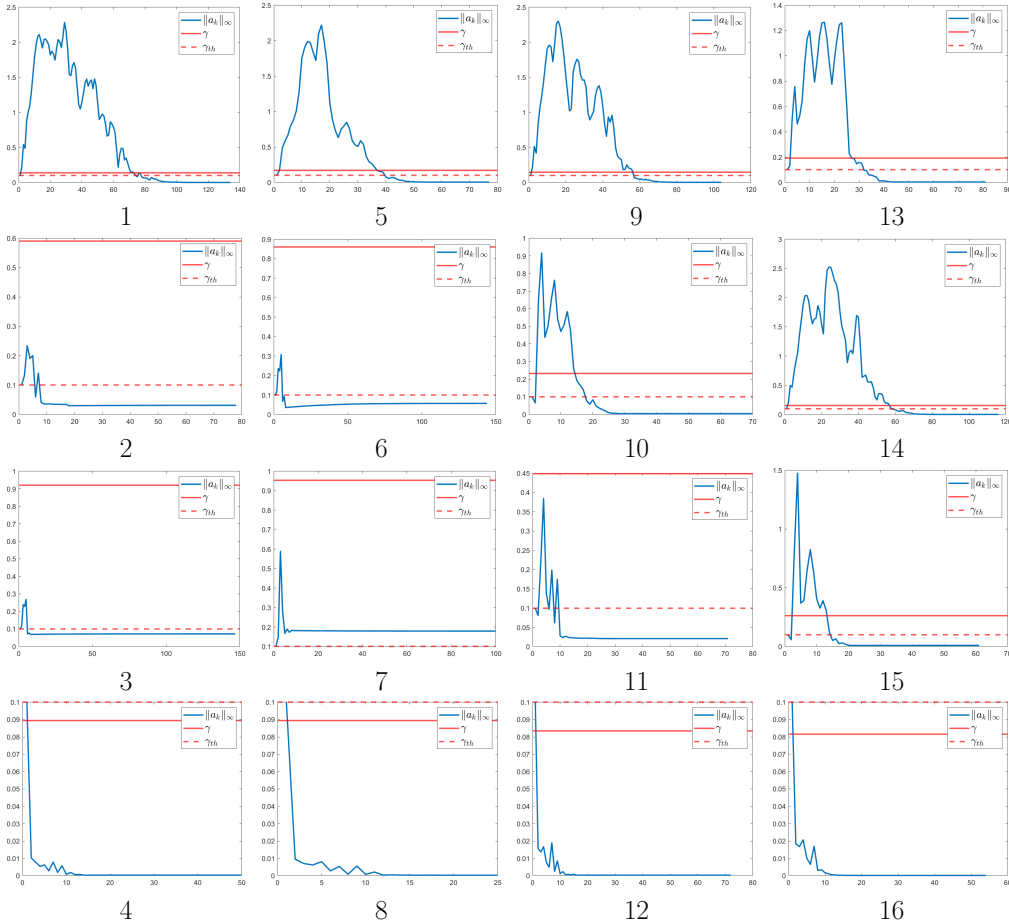


Figure 6: Maximum absolute coefficient value $\|a^\kappa\|_\infty$ throughout the iterations $\kappa = 1, 2, \dots$ (blue), mean value of the γ -map (solid red) and threshold γ_{th} (dashed red) for some selected patches.

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. In the third example, we compare our method with the dictionary-based reconstruction approach for OCT images proposed in previous work [52]. Finally, in the fourth example, we will show how our reconstruction is well-suited to improve the quality of post-processing tasks such as 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. 7, 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 γ -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 γ -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 γ -maps in Fig. 7 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 γ 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

Patch	BFGS		CBFGS			NNC-BFGS		
	nit	time	nit	time	cau	nit	time	γ
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	50	6.41	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	25	2.58	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	72	9.5	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	54	6.62	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 γ for the current patch.

have an excessive impact on the quality of the reconstructed HR OCT image. Complementarily, low γ values in the bottom part of the γ -maps, force the non-convexity of the penalty in (P_y) , thus enforcing strong sparsity of solution, which leads to a predominantly zero reconstruction.

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 [53]. 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:

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

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

The regularisation parameter λ is globally fixed throughout the image. In this manner the effect of the penalty is homogeneous throughout the image according to the given λ 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 λ 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 γ_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 $(P_y - \text{MCP})$ can be efficiently solved by means of proximal gradient strategy, as explained in [45].

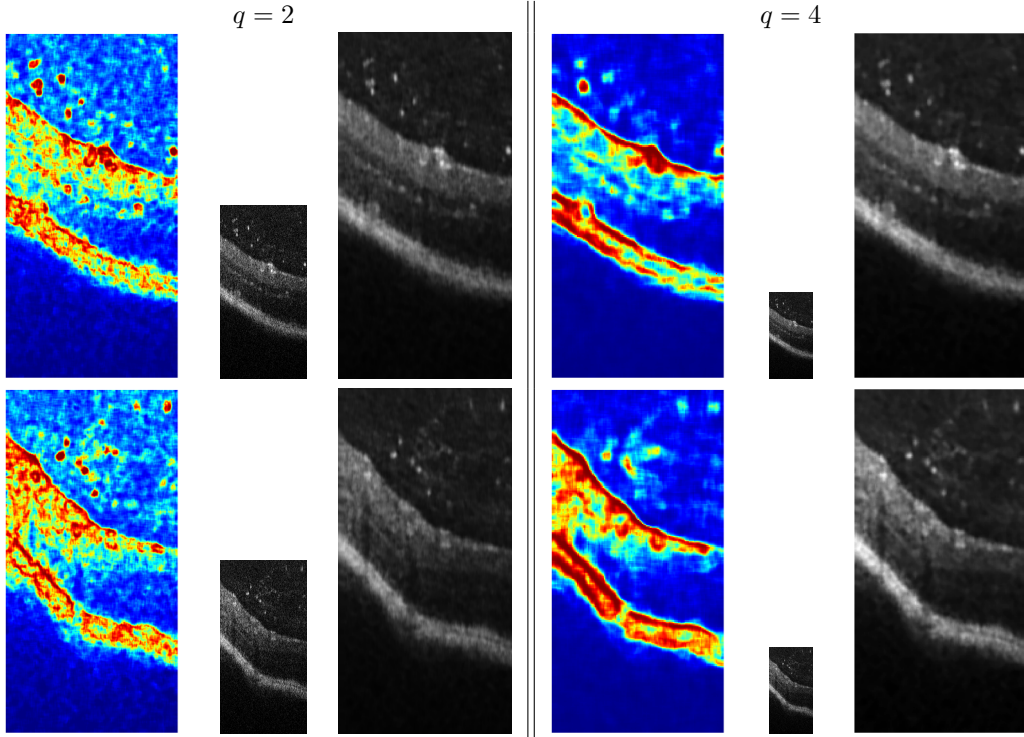


Figure 7: Reconstruction for OCT3 (first row) and OCT4 (second row) with scaling factor $q = 2$ (left panel) and $q = 4$ (right panel). Left panel, from left to right: γ -map (512×256), LR data (512×256), HR reconstruction (1024×512). Right panel, from left to right: γ -map (256×128), LR data (256×128), HR reconstruction (1024×512).

The reconstructions obtained by means of these different regularisers are compared on OCT1 image (Fig. 8, first row) and OCT2 image (Fig. 8, 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 central and in the right panels are obtained by running Algorithm 1 with optimisation problems $(P_y - \ell_1)$ and $(P_y - \text{MCP})$, respectively. For both ℓ_1 and MCP regularisers, we report the reconstructed OCT images obtained using a global high and low λ value. For high values of λ 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 λ an accurate reconstruction of the textured layers is obtained, which, however, is unable to remove the background noise at the same time.

6.3. Comparison with a dictionary-based reconstruction framework

We now compare the reconstruction obtained by the proposed approach with the dictionary-based reconstruction approach for OCT images proposed in [52]. There, an upscaled version of the LR image is obtained via a bicubic interpolation. Then, as a second step, HR patches are extracted from the interpolated image and reconstructed as a linear combination of the elements of the dictionary. This is obtained by solving for each patch an optimisation problem that is similar to (P_y) , but without downsampling operator \mathcal{S}_q and with a non-separable sparsity-promoting generalised MCP term. The final solution is then obtained by reassembling the reconstructed patches and taking the mean of the overlapping values upon the same pixel. We carried out a reconstruction of factor $q = 4$ on images OCT1, OCT2, OCT3, and OCT4, using the dictionary D built as above. For the remaining hyperparameters of the model (*e.g.* the regularisation parameter) we kept the default settings proposed by the authors. The results are reported in Fig. 9. It can be noticed that the last group of OCT images contain saturated values which make their inspection more challenging. In addition, the background noise in the upper part of the images is enhanced

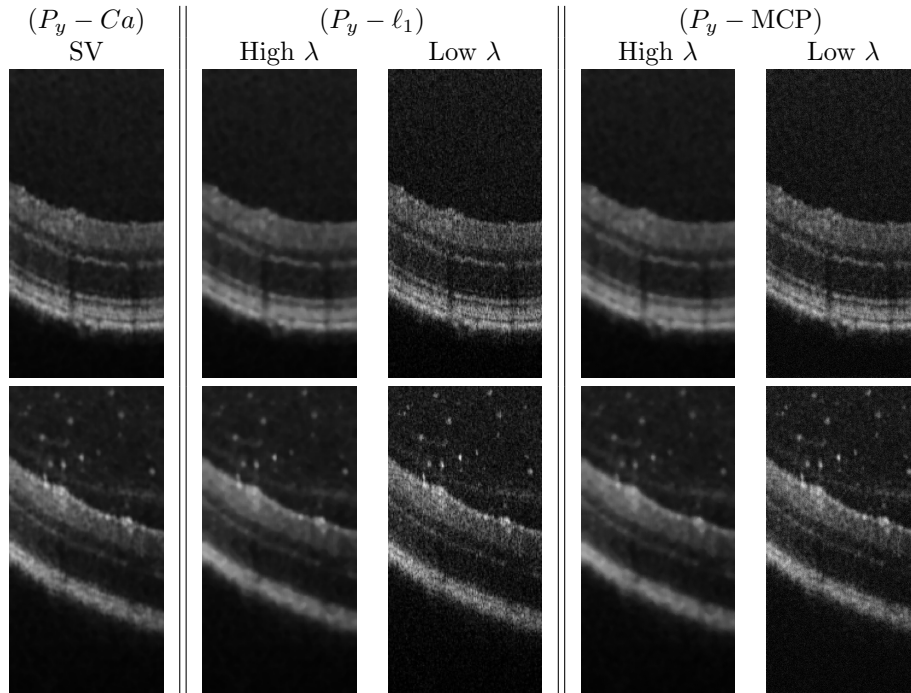


Figure 8: Effects of different regularisers in (P_y) on the OCT image reconstruction.

and turned into artefacts, which might hinder a precise identification of the cells suspended in the vitreous.

6.4. 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 [50, 9]. Here, we illustrate how the use of the proposed reconstruction model may improve the quality in OCT image segmentation.

Recalling Section 4, one might think of obtaining a first partitioning result directly from the analysis of the γ -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. 10, where on the left we report the $\gamma_{i,j}$ values estimated by means of the procedure detailed in Section 4, while on the right we illustrate its segmentation by means of a simple k -means-based partitioning approach, where the value $k = 4$ was selected so as to represent different regions in the image. As it can be noticed, such segmentation properly separates the main DS regions in the image and spots some 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. In a similar way, in Fig. 11 first left panel, we show on OCT2, OCT3 and OCT4 how the direct segmentation of the LR data does not yield satisfactory results either. For the segmentation step we exploited the already mentioned k -means with $k = 4$ and the multiphase Chan-Vese (CV) method [49].

Better results can be obtained by performing image segmentation on the HR data, yielding in particular higher precision in detecting the layered region within the retina and the portions of tissue suspended in the vitreous. In particular, we show how the proposed reconstruction method allows to obtain interesting results in combination with different segmentation models (k -means and CV). In Fig. 11 second panel, we show this improvement with respect to the LR case (first left panel). The application of the proposed adaptive SR preliminary process (A-OCT-SR) allowed us to achieve remarkable results in separating the layers of the retinal structure, with simultaneous background noise removal. This can be seen in the fact that on these images, both k -means and CV yielded an homogeneous labelling of the layers and of the elements suspended in the vitreous. In this example we compare the segmentations obtained on our HR results with the ones obtained

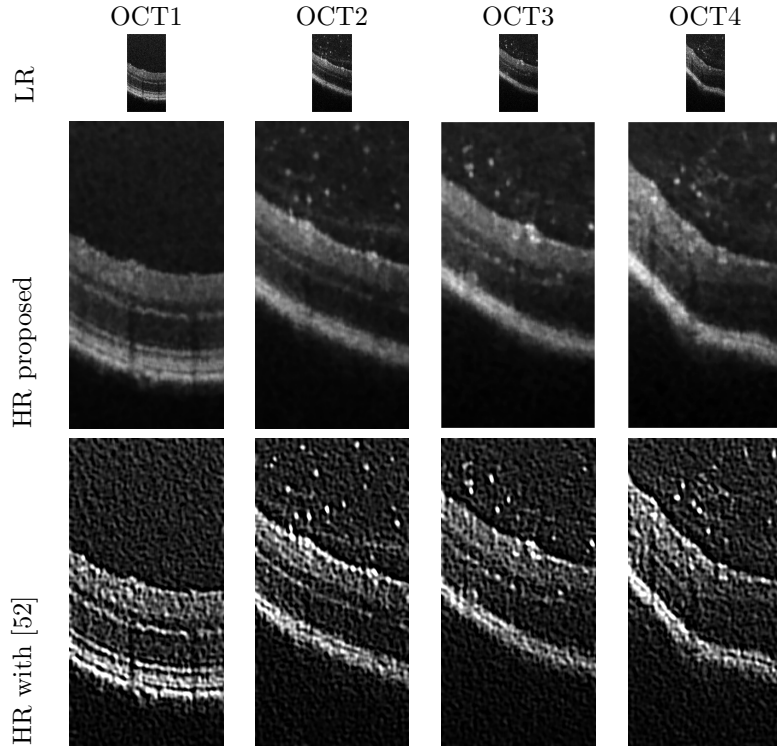


Figure 9: Comparison on OCT1, OCT2, OCT3 and OCT4 images of LR data (first row), of the proposed reconstruction (second row) with the reconstruction obtained using [52] (third row).

on the HR images produced by a Compressed-Sensing-based variational SR approach (CSSR) proposed in [29]. It can be noticed that these images still reproduce the noisy, pixel-wise structure of their LR counterparts and the corresponding k -means segmentation is fragmented, while, on the other hand, the CV results are less accurate in fine-structure detection. Finally, in the last panel of Fig. 11, we also compare with a joint variational Super-Resolution-and-Segmentation approach (J-SR), introduced in [35], where the segmentation is directly obtained by promoting an ℓ_0 -sparsity piece-wise constant reconstruction. This joint segmentation approach could be optimal in this context but it completely destroys the texture of the OCT image. In addition, this model is difficult to tune as it is highly parameter dependent and it thus requires the operator's expertise in order to be used.

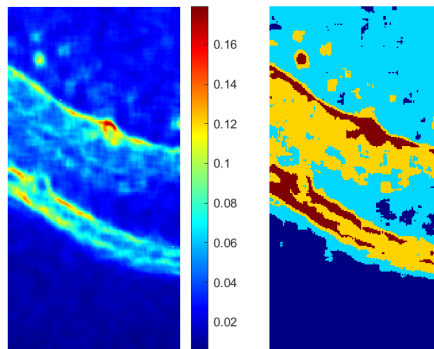


Figure 10: Segmentation: $\gamma_{i,j}$ values of a LR OCT image (left) and their k -means partitioning with $k = 4$ (right)

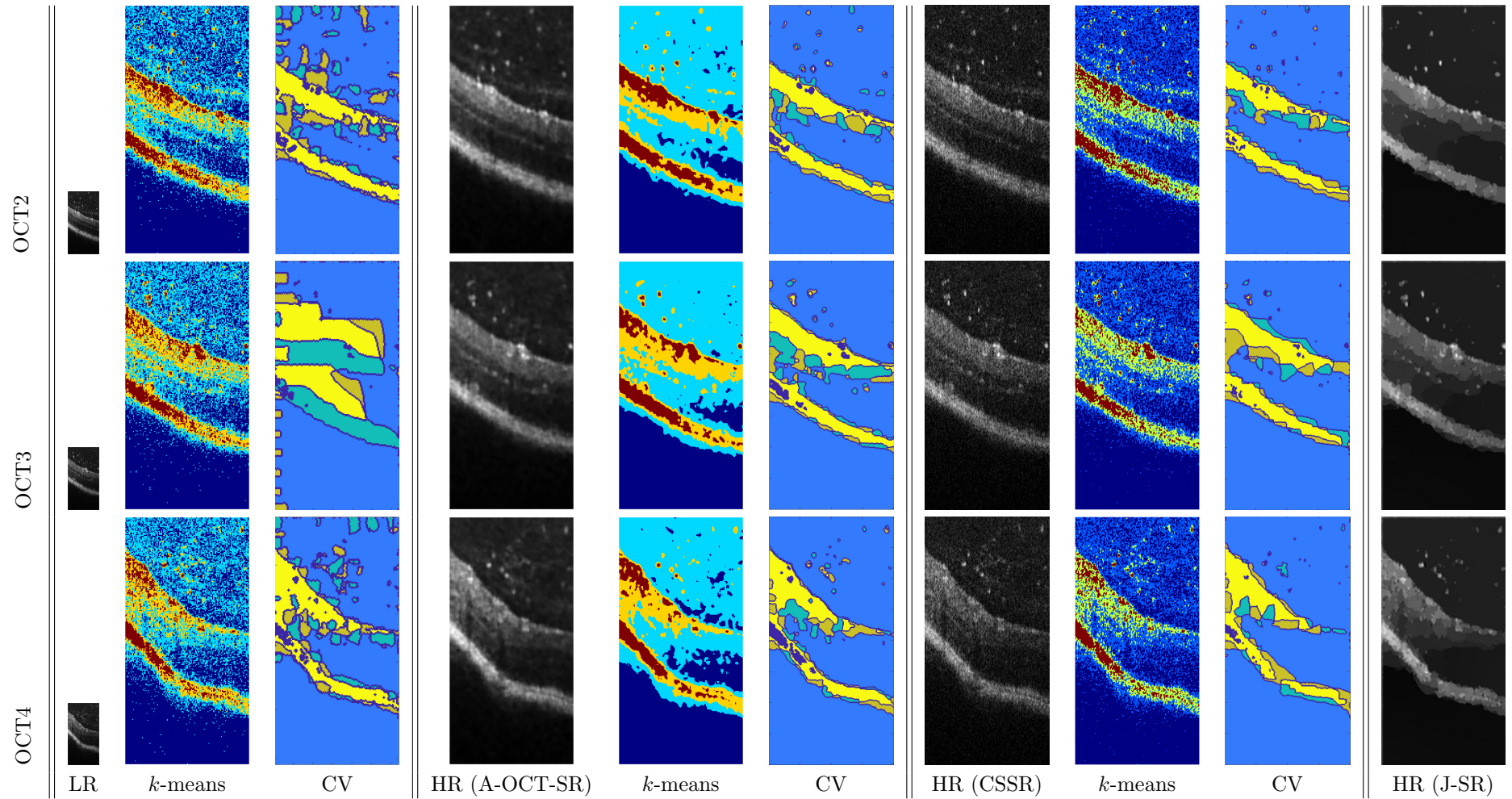


Figure 11: Reconstruction and segmentation of OCT2, OCT3 and OCT4, from left to right: LR image, Adaptive OCT-SR reconstruction, CSSR reconstruction, Joint SR and segmentation.

7. Conclusions

We proposed an adaptive, smooth and non-convex, variational approach for the reconstruction of high-resolution OCT data from low-resolution measurements using a sparse representation framework. Motivated by statistical considerations, a Cauchy-based penalty was used to promote sparsity. Differently from more standard models, traditionally used to approximate the ℓ_0 pseudo-norm, the Cauchy penalty depends on a scale parameter whose size determines not only the regularisation strength, but also its shape. A patch-adaptive parameter estimation strategy based on Maximum Likelihood is proposed. It allows to take into account the heterogeneous content observed in real OCT data. The strategy was shown to produce an adaptive map of parameters, which self-tunes the regularisation strength depending on the type of local information observed. In order to exploit the overall smoothness of the model considered, which combines the Cauchy penalty with an ℓ_2 data fit, we opted for a variant of the well-known BFGS algorithm where a cautious update of the descent direction is used to adapt to the different convex vs. non-convex regime. Several numerical tests were reported showing gains in computational efficiency. We validated the overall adaptive OCT-SR algorithm on different real OCT images. We showed the advantages of having an adaptive regularisation model with respect to a fixed one, and compared our results with the ones obtained by means of popular non-smooth convex/non-convex regularisers frequently employed to enforce sparsity. Finally, we showed how the reconstructed data can be successfully employed for image segmentation purposes by means of standard algorithms. Our future work will focus on further investigating the advantages linked to the use of the Cauchy penalty, in order to demonstrate the benefits of employing smooth, albeit non-convex, loss functions in solving various inverse problems. On the one hand, we aim to demonstrate the possibility of using it as a component within variational autoencoders, while on the other hand we hope to be able to illustrate its benefits when techniques relying on deep unfolding (see e.g. [34, 6]) are employed.

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 “Attractivité du territoire” by UCA IDEX JEDI. Scrivanti acknowledges the financial support from the the ITN-ETN project TraDE-OPT funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861137 (this work represents only the authors’ view and the European Commission is not responsible for any use that may be made of the information it contains). 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 α -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] S. Aja-Fernández, T. Pieciak, and G. Vegas-Sánchez-Ferrero. Spatially variant noise estimation in mri: A homomorphic approach. *Medical Image Analysis*, 20:184 – 197, 2015.

- [6] A. Balatsoukas-Stimming and C. Studer. Deep unfolding for communications systems: A survey and some new directions. In *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, pages 266–271. IEEE, 2019.
- [7] L. R. Borges, M. A. C. Brochi, Z. Xu, A. Foi, M. A. C. Vieira, and P. M. Azevedo-Marques. Noise modeling and variance stabilization of a computed radiography (CR) mammography system subject to fixed-pattern noise. *Physics in Medicine & Biology*, 65(22):225035, nov 2020.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001.
- [15] 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.
- [16] 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.
- [17] L. Dong, J. Zhou, and Y. Y. Tang. Content-adaptive noise estimation for color images with cross-channel noise modeling. *IEEE Transactions on Image Processing*, 28(8):4161–4176, 2019.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] M. Huska, D. Lazzaro, S. Morigi, A. Samorè, and G. Scivanti. Spatially-adaptive variational reconstructions for linear inverse electrical impedance tomography. *Journal of Scientific Computing*, 84(3):46, 2020.

- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] D. Lazzaro, S. Morigi, P. Melpignano, E. Loli Piccolomini, and L. Benini. Image enhancement variational methods for enabling strong cost reduction in oled-based point-of-care immunofluorescent diagnostic systems. *International Journal for Numerical Methods in Biomedical Engineering*, 34(3):e2932, 2018. e2932 cnm.2932.
- [30] 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.
- [31] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:674–692, July 1989.
- [32] P. McCullagh and N. G. Polson. Statistical sparsity. *Biometrika*, 105(4):797–814, 10 2018.
- [33] O. V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 53(1):64–78, 2006.
- [34] V. Monga, Y. Li, and Y. C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.
- [35] D. Mylonopoulos, P. Cascarano, L. Calatroni, and E. L. Piccolomini. Constrained and Unconstrained Inverse Potts Modelling for Joint Image Super-Resolution and Segmentation. *Image Processing On Line*, 12:92–110, 2022. <https://doi.org/10.5201/ipo1.2022.393>.
- [36] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.
- [37] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley and Sons, New York, 1995.
- [38] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.
- [39] 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.

- [40] M. Pereyra and S. McLaughlin. Fast unsupervised bayesian image segmentation with adaptive spatial regularisation. *IEEE Trans. Image Proc.*, 26:2577–2587, 2017.
- [41] M. J. D. Powell. On the Convergence of the Variable Metric Algorithm. *IMA Journal of Applied Mathematics*, 7(1):21–36, 02 1971.
- [42] M. Pragliola, L. Calatroni, A. Lanza, and F. Sgallari. On and beyond Total Variation regularisation in imaging: the role of space variance, 2021. arXiv preprint at: <https://arxiv.org/abs/2104.03650>, to appear in SIAM review.
- [43] P. Puvanathan 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.
- [44] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.
- [45] 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.
- [46] E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for ℓ_2 - ℓ_0 minimization. *SIAM Journal on Optimization*, 27(3):2034–2060, 2017.
- [47] 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.
- [48] M. Unser and P. D. Tafti. *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2014.
- [49] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International journal of computer vision*, 50(3):271–293, 2002.
- [50] Q. Wang, R. Zheng, and A. Achim. Super-resolution in optical coherence tomography. In *40th Annual IEEE EMBS Conference*, pages 1–4, 2018.
- [51] 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.
- [52] 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.
- [53] C.-H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.*, 38(2):894–942, 04 2010.
- [54] 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.