Non-convex Super-resolution of OCT images via sparse representation

Gabriele Scrivanti, Luca Calatroni, Serena Morigi, Lindsay Nicholson, Alin Achim

HAL Id: hal-02978016

https://hal.science/hal-02978016

Submitted on 29 Nov 2021

NON-CONVEX SUPER-RESOLUTION OF OCT IMAGES VIA SPARSE REPRESENTATION

Gabriele Scrivanti[#†]   *Luca Calatroni*[⋆]   *Serena Morigi*[†]   *Lindsay Nicholson*[‡]   *Alin Achim*[‡⋆]

[#] Center for Visual Computing - OPIS Inria group - CentraleSupélec - University Paris-Saclay, France
[†] Department of Mathematics, University of Bologna, Italy
[⋆] Laboratoire I3S - CNRS - UCA - Morpheme Inria group - Sophia-Antipolis, France
[‡] University of Bristol, UK

ABSTRACT

We propose a non-convex variational model for the super-resolution of Optical Coherence Tomography (OCT) images of the murine eye, by enforcing sparsity with respect to suitable dictionaries learnt from high-resolution OCT data. The statistical characteristics of OCT images motivate the use of α-stable distributions for learning dictionaries, by considering the non-Gaussian case, $\alpha = 1$. The sparsity-promoting cost function relies on a non-convex penalty - Cauchy-based or Minimax Concave Penalty (MCP) - which makes the problem particularly challenging. We propose an efficient algorithm for minimizing the function based on the forward-backward splitting strategy which guarantees at each iteration the existence and uniqueness of the proximal point. Comparisons with standard convex ℓ_1-based reconstructions show the better performance of non-convex models, especially in view of further OCT image analysis.

Index Terms— Optical Coherence Tomography, Super-Resolution, Non-Convex Regularisation, Sparse Representation.

1. INTRODUCTION

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) so 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, Alzheimer's disease, Parkinson disease, or Glaucoma [1]. However, the poor spatial resolution and the multiplicative nature of the (speckle) noise observed in OCT images often limits 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 [1], often based on accurate (and often manual) layer segmentation. In this paper, we propose a non-convex variational framework for the super resolution

of real murine OCT images based on sparse representations with respect to pre-computed high-resolution OCT dictionaries. The SR problem is formulated as the inverse problem of retrieving the original high-resolution (HR) image from a given low-resolution (LR) one where we account also for the presence of blur and background noise. We further assume that the entries of the desired HR image are symmetric-α-stable random variables and thus the OCT image can be well represented by only a few atoms of the given HR dictionary learned by α-stable distributions following [2]. In our framework, the super-resolution is performed patch-wise as part of the reconstruction procedure as presented in [3], differently from the alternative approach presented in [1] for SR of OCT images, where the upscaling is obtained via an interpolation step which is preliminary to the patch extraction. We enforce sparsity by means of a non-convex regularisation term, and, in particular, we consider the separable Cauchy-based penalty and the MCP, which both depend on a scalar positive parameter which modulates the non-convex behaviour. Following [4], we efficiently solve the non-convex problem by means of a forward-backward splitting algorithm [5] where the existence and uniqueness of the proximal point are guaranteed by suitable conditions introduced on the model and algorithmic parameters. This is different from the recent Convex-Non-Convex (CNC) approaches [6] where conditions on the overall convexity of the total objective functional are derived, while preserving the non-convex behaviour of the regularisation term. The proposed strategy, working in a pure non-convex regime, guarantees, on the other hand, convergence to a stationary point. We show quantitative and qualitative comparisons with reconstructions obtained by means of the ℓ_1-norm based model and different dictionaries.

2. SR VIA SPARSE REPRESENTATION

The task of recovering a 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 modeled mathematically as an inverse image reconstruction problem whose ill-posedness, as described in [3], can be overcome by representing X in a sparse way with respect to a given (over-complete) dictionary. The overall problem can be thus mod-

elled by introducing the following constraints.

1) *Reconstruction Constraint*: the input image Y is linked to the desired HR image X via the model

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

where $\mathcal{H} : \mathbb{R}^{r_h \times c_h} \rightarrow \mathbb{R}^{r_h \times c_h}$ is the blur operator corresponding to the point spread function (PSF) of the optical acquisition system considered, $\mathcal{S}_q : \mathbb{R}^{r_h \times c_h} \rightarrow \mathbb{R}^{r_h/q \times c_h/q}$ is a downsampling operator defined in terms of a factor $q \in \mathbb{N}$ which maps the blurred image $\mathcal{H}(X) \in \mathbb{R}^{r_h \times c_h}$ into a coarser grid by averaging. As any imaging technique that is based on detection of coherent waves, OCT images are subject to the presence of speckle noise [7], thus $\mathcal{N}(\cdot)$ stands for a multiplicative (speckle) noise degradation process, which is not considered in this work, while $\eta \in \mathbb{R}^{r_h/q \times c_h/q}$ denotes an additive component which represents white Gaussian background noise.

2) *Sparsity Constraint*: We 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 overcomplete dictionary $\mathcal{D} \in \mathbb{R}^{n_p \times n_d}$ which has been previously learned in terms of HR training images. Using a vectorised notation for any patch $x \in \mathbb{R}^{n_p}$, this assumption translates in:

$$x \approx \mathcal{D}a \quad \text{for} \quad a \in \mathbb{R}^{n_d} \quad \text{with} \quad \|a\|_0 << n_d. \qquad \text{(C2)}$$

To learn the dictionary $\mathcal{D}$, we consider the approach presented in [2] where general α-Stable distributions are used as prior PDFs. In particular, we set the parameter $\alpha = 1$ so that the underlying distribution is assumed to be Cauchy.

We combine (C1) with (C2) to compute the sparse coefficient vector a for each given LR square patch $y \in \mathbb{R}^{n_p/q^2}$ extracted from the observed image Y, thus getting the following optimisation problem

$$a^* \in \underset{a \in \mathbb{R}^{n_d}}{\arg\min} \left\{ \frac{1}{2} \|y - \mathcal{S}_q(\mathcal{H}(\mathcal{D}a))\|_2^2 + \lambda \sum_{i=1}^{n_d} \phi_\gamma(a_i) \right\}, \quad \text{(P)}$$

where ϕ_γ is a numerically tractable version of the ℓ_0 pseudo-norm in (C2) promoting sparsity and depending on a parameter $\gamma > 0$, and $\lambda > 0$ is a regularisation parameter.

The solution a^* of the optimisation problem (P) is the sparse coefficient vector representing the HR patch $x \in \mathbb{R}^{n_p}$ under the over-complete dictionary $\mathcal{D}$.

3. SPARSE NON-CONVEX REGULARISATIONS

We describe in this section two popular sparsity-promoting, non-convex and parametric penalty functions $\phi_\gamma : \mathbb{R} \rightarrow \mathbb{R}^+$ in (P) which are considered in the proposed SR framework. Among the non-convex sparsity-promoting regularizers characterized by tunable degree of non-convexity, the MCP is considered one of the most interesting and effective penalties [6]. As a natural alternative to the MCP we propose the Cauchy-based penalty, assuming that the desired solution vector a^*
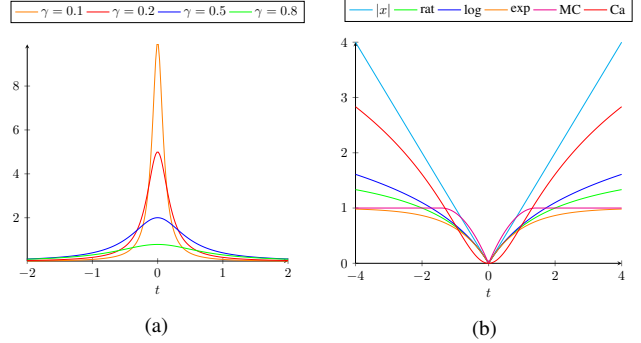

Fig. 1: (a) Cauchy PDF for varying $\gamma > 0$. (b) Plots of different penalty functions relaxations of the ℓ_0 pseudo-norm.

is distributed as an n_d-dimensional Cauchy distribution with parameter γ; it is therefore easy to derive an instance of (P) following, e.g., [4].

The Cauchy distribution belongs to the family of α-stable distributions (with $\alpha = 1$) which are heavy-tailed and have been exploited in imaging applications, see *e.g.* [2]. Differently from other probability distributions in the family, in the case of Cauchy distribution there exists a closed-form probability density function whose negative log-likelihood for $t \in \mathbb{R}$ reads:

$$\phi_\gamma(t) = \phi_\gamma^C(t) = \log\left(\frac{\gamma^2 + t^2}{\gamma}\right). \qquad (1)$$

As illustrated in Figure 1a the parameter γ controls the spread of the distribution: the smaller γ, the narrower and more peaked the shape of the distribution. We remark that the function ϕ_γ in (1) is non-convex except for a small and limited interval around the origin and increases unbounded at a logarithmic rate. In Figure 1b we report the behaviour of the Cauchy penalty when used as a relaxations of the ℓ_0 pseudo-norm in comparisons with other classical choices as ℓ_1-norm, rat, log, exp, MCP, see [6] for details. Finally, we remark that the only convex function is the ℓ_1-norm function, which corresponds to $\phi_\gamma(t) = |t|$.

The Minmax Concave Penalty introduced in [8] has been successfully applied to several signal and image recovery problems. The function ϕ_γ in this case reads:

$$\phi_\gamma(t) = \phi_\gamma^{MCP}(t) = \begin{cases} -\frac{1}{2\gamma}t^2 + \sqrt{\frac{2}{\gamma}}t & \text{if } |t| < \sqrt{2\gamma}, \\ 1 & \text{if } |t| \geq \sqrt{2\gamma}, \end{cases} \qquad (2)$$

where $\gamma > 0$ modulates the concavity of the regulariser.

4. FORWARD-BACKWARD ALGORITHM

For every patch, we solve the non-convex problem (P) by means of a Forward-Bacwkard (FB) splitting algorithm, then we recombine the patches together via sliding averaging. Algorithm 1 illustrates the pseudocode where D, and S_q denote suitably resized matrices computed by the operators $\mathcal{D}$ and

$\mathcal{S}_q$, respectively. The blur matrix H is related to the PSF of the OCT scan which is estimated as described in [7].

The convergence of the FB algorithm is subordinated to a sufficiently small step size μ in the forward step [F], and the existence of a unique minimizer of the backward step [B]. Regarding μ, according to [9], the range of step size values that guarantees convergence is $(0, 1/L)$, where $0 < L = \|(S_q H D)^T S_q H D\|$ denotes the Lipschitz constant of the quadratic fidelity term. In order to have a uniquely defined proximal point at each backward step [B] we look for a parameter value γ such that:

$$\text{prox}_{\lambda\mu\Phi_\gamma}(x) = \underset{u \in \mathbb{R}^{n_d}}{\arg\min} \left\{ J(u) := \frac{\|x - u\|_2^2}{2\lambda\mu} + \sum_{i=1}^{n_d} \phi_\gamma(u_i) \right\}$$

is a singleton (i.e. $J(\cdot)$ is strongly convex). By the separability of the penalty $\Phi_\gamma(\cdot) := \sum_{i=1}^{n_d} \phi_\gamma(\cdot)$ we first observe that the proximal point can be computed component-wise on each i-th component, $i = 1, \ldots, n_d$. Then, we follow the strategy in [4, Theorem 2] and choose, for both the Cauchy (1) and the MCP (2) penalties, a parameter γ such that J is strictly convex, which guarantees the uniqueness of proximal point. Balancing the positive second derivatives in the first term (fidelity) against the negative second derivatives in the regularization term of J, we find that the strict convexity is guaranteed when the following condition holds for $\tau > 1$:

$$\gamma = \tau\bar{\gamma}, \quad \bar{\gamma} := \begin{cases} \frac{\sqrt{\lambda\mu}}{2} & \text{if} \quad \phi_\gamma = \phi_\gamma^C, \\ \lambda\mu & \text{if} \quad \phi_\gamma = \phi_\gamma^{MCP}. \end{cases} \quad (3)$$

Note that, in both cases, a closed form expression of the proximal points can be easily computed by simply looking for solutions of the corresponding optimality conditions. In the case of Cauchy regularisation the solution can be computed using Cardano's method for solving the cubic optimality condition [10], while for MCP a closed formula is proposed in [8]. The extension of this algorithmic strategy to the design of converging FB algorithms for general non-convex parametric regularisers ϕ_γ is left for future research. We remark that (3) does not affect the overall non-convexity of the original problem (P). Hence, only convergence to a stationary point can be guaranteed.

For each given LR patch y, once the sparse vector a is obtained, the reconstructed HR patch x is computed by enforcing the condition (C2). Note that by definition, the HR patch computations are independent between each other and as such they can be parallelised. As a final step, the reconstructed SR image is generated by aggregating all the HR patches together and locally averaging according to the number of overlapping patches at each location. This was shown in [1] to produce reconstruction results with less artefacts.

Algorithm 1 FB splitting for patch-based sparse OCT SR

Input: Y, D, S_q, $\mu \in (0, 1/L)$, $\lambda > 0$, $\gamma \in (\bar{\gamma}, +\infty)$ **Output:** X^* % *Super-resolved OCT image*
Estimate PSF as in [7] and get H Extract overlapping patches of size $\frac{\sqrt{n_p}}{q} \times \frac{\sqrt{n_p}}{q}$ **For** each patch y:
    $a_0 = 0$
    For $k = 0, 1, \ldots$ **do:**
        $b_{k+1} = a_k - \mu D^T H^T S_q^T (y - S_q H D a_k)$ [F]
        $a_{k+1} = \text{prox}_{\lambda\mu\Phi_\gamma}(b_{k+1})$ [B]
    until convergence
    $x = Da$ % *Generate the HR patch*
end
X^* is obtained by sliding average of the HR patches

---

5. NUMERICAL RESULTS

In this section we discuss the performance of the patch-based sparse OCT SR Algorithm 1 when applied to the case of Cauchy, MCP and ℓ_1 regularisation and for different choices of the dictionary D. The quantitative assessment of the SR performance for all tests is based on the comparison between the obtained SR image $X^* \in \mathbb{R}^{r_h \times c_h}$ and a reference HR version $X^0 \in \mathbb{R}^{r_h \times c_h}$ of the input image in terms of Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). We remark that these comparisons are useful to assess only the SR capabilities of the different approaches considered. Due to the lack of reference ground-truth images, in practical situations, however, a qualitative visual evaluation performed by a practitioner is still required to assess the computed results in view of further analysis (typically, layer segmentation).

Dictionaries and parameters. We considered 4 dictionaries denoted by $\mathcal{D}(n_p, n_d)$ which vary according to the number of atoms $n_d \in \{300, 600\}$ and the dimension of each atom $n_p \in \{64, 256\}$. The reason behind this choice is to assess how the features of the dictionary affect the resulting reconstruction and to identify which combination is the most suitable for the image analysis task (typically, segmentation or detection) one aims to perform next. To build the dictionaries, we used a sample of 60 noise-free HR OCT images as a training set and applied the SparseDT approach described in [2] with the assumption that the underlying data distribution corresponds exactly to the Cauchy Distribution. We consider a SR model with magnification factor $q = 4$, so that $Y \in \mathbb{R}^{256 \times 128}$ and $X^* \in \mathbb{R}^{1024 \times 512}$. The FB iterations stop as soon as the tolerance 10^{-5} is satisfied or whenever the maximum number of 300 iterations is reached. Recalling the condition (3), we set $\tau = 1.01$, we let λ vary in the range $[10^{-7}, 1]$. Table 1 reports the PSNR and SSIM values of the solutions obtained by the possible combinations of the 4 dictionaries and the 3 penalties for two test images which we refer to as OCT1 and OCT2. We notice that the best quantitative results are obtained in correspondence of the

		$\mathcal{D}_{(256,300)}$	$\mathcal{D}_{(64,300)}$	$\mathcal{D}_{(256,600)}$	$\mathcal{D}_{(64,600)}$
OCT1	Cauchy	19.923	20.079	19.059	**20.395**
	MCP	19.871	20.124	20.386	20.386
	ℓ_1	19.928	20.288	19.259	20.376
OCT2	Cauchy	19.560	19.893	18.694	**20.200**
	MCP	19.512	19.747	17.657	20.148
	ℓ_1	19.564	20.001	19.300	20.185

Table 1: PSNR values for the reconstruction of the two test images.
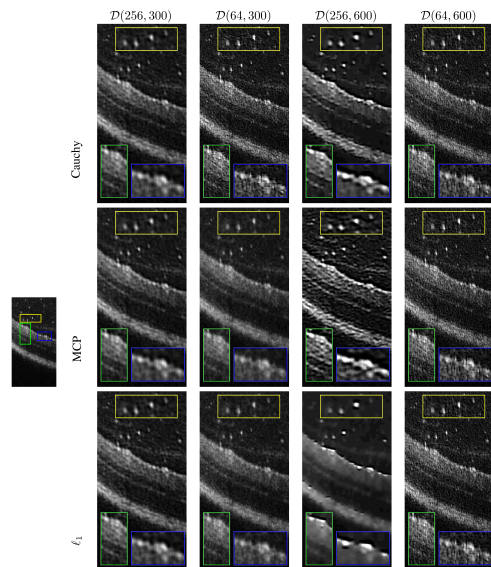
Fig. 2: OCT1 Super-Resolution reconstructions ($q = 4$).

Fig. 3: Segmentation with k-means, $k = 3$, before (LR) and after SR for different regularisations: Cauchy, MCP, ℓ_1.

dictionaries with $n_p = 64$, while, on the contrary, the dictionaries with $n_p = 256$ generally lead to lower quantitative values, but more visually satisfying results, especially in view of further processing of the image such as segmentation and detection. We report the results in Fig. 2 where we frame some details in yellow, blue and green showing respectively some cells suspended in the vitreous, the separation between the vitreous and the retina and finally a portion of the upper layer region. We remark that the combination of $\mathcal{D}(256, 600)$ and a non-convex penalty achieves remarkable results in separating the layers of the retinal structure, with Cauchy being the one better removing background noise at the cost of a slight smoothing in the layer region, which can help layer classification analysis.

We support our discussion reporting in Fig. 3 the k-means segmentation of the reconstructed images for OCT1 and OCT2 with $k = 3$, which corresponds to the regions of interest: background, layers and textured layers.

6. CONCLUSIONS

We considered a non-convex variational model for patch-based SR of highly-degraded real murine OCT images promoting sparsity w.r.t. to a heavy-tailed Cauchy dictionary. We enforce sparsity by means of separable non-convex parametric regularisations and assess the performance of the SR model both quantitatively and visually. By imposing a condition on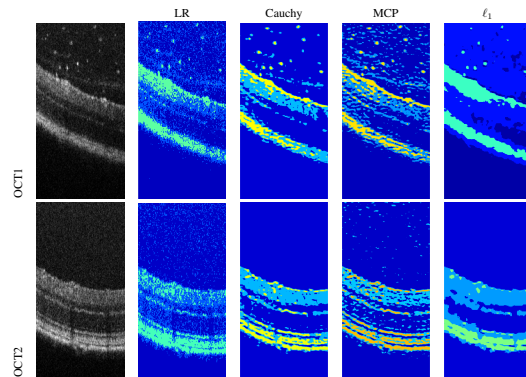 the model/algorithmic parameters preserving the overall non-convexity of the composite functional, we guarantee the existence of a unique proximal point. Possible extensions involve generalisation to other regularisers and application to the analysis of OCT data.

9. REFERENCES

[1] Q. Wang, R. Zheng, and A. Achim, "Super-resolution in optical coherence tomography," in *40th Annual IEEE EMBC*, 2018, pp. 1–4.

[2] P. Pad, F. Salehi, E. Celis, P. Thiran, and M. Unser, "Dictionary learning based on sparse distribution tomography," 2017, vol. 70 of *Proceedings of Machine Learning Research*, pp. 2731–2740, PMLR.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19(11), 2010.

[4] O. Karakuş, P. Mayo, and A. Achim, "Convergence guarantees for non-convex optimisation with cauchy-based penalties," *IEEE Trans. on Signal Proces.*, vol. 68, pp. 6159–6170, 2020.

[5] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul., 4(4)*, pp. 1168–1200, 2005.

[6] A. Lanza, S. Morigi, and F. Sgallari, "Convex image denoising via non-convex regularization with parameter selection," *J. Math. Imaging Vis.*, 2016.

[7] O.V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 53(1), 2006.

[8] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 04 2010.

[9] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Math. Program., Series A*, vol. 137, no. 1, pp. 91–124, Aug. 2011.

[10] T. Wan, N. Canagarajah, and A. Achim, "Segmentation of noisy colour images using cauchy distribution in the complex wavelet domain," *IET Image Processing*, vol. 5, no. 2, pp. 159–170, 2011.