Efficient ℓ Gradient-Based Super-Resolution for Simplified Image Segmentation

(Article begins on next page)

10 April 2024

Efficient ℓ^0 gradient-based super-resolution for simplified image segmentation

Pasquale Cascarano, Luca Calatroni, Elena Loli Piccolomini

Abstract—We consider a variational model for single-image super-resolution based on the assumption that the gradient of the target image is sparse. We enforce this assumption by considering both an isotropic and an anisotropic ℓ^0 regularisation on the image gradient combined with a quadratic data fidelity, similarly as studied in [1] for signal recovery problems. For the numerical realisation of the model, we propose a novel efficient ADMM splitting algorithm whose substeps solutions are computed efficiently by means of hard-thresholding and standard conjugate-gradient solvers. We test our model on highly-degraded synthetic and real-world data and quantitatively compare our results with several sparsity-promoting variational approaches as well as with state-of-the-art deep-learning techniques. Our experiments show that thanks to the ℓ^0 smoothing on the gradient, the super-resolved images can be used to improve the accuracy of standard segmentation algorithms for applications like QR codes and cell detection and land-cover classification problems.

Index Terms—Single-image super-resolution, ℓ^0-gradient regularization, image segmentation, ADMM.

I. Introduction

The task of single-image Super-Resolution (SR) consists in improving the spatial resolution of an observed Low-Resolution (LR) imaging data so as to obtain a High-Resolution (HR) version which, typically, can be used as a reference for subsequent analysis. Image resolution is limited in many applications due to the optical characteristics and the physical limitations of the acquisition devices. Some standard examples are biomedical and astronomic imaging where, due to light aberration phenomena, close objects (molecules, stars. . .) on LR images cannot be correctly distinguished/detected, see, e.g. [2], [3]. SR techniques are often employed also in image recognition problems. This is the case, for instance, of QR code recognition where images are often captured by scanning tools (e.g. cell-phones) from relatively large distances which may affect the accuracy of the recognition [4]. Analogously, in remote sensing applications such as land-cover classification, multi- and hyperspectral measurements often suffer from poor spatial resolution, which may limit significantly the classification precision [5], [6].

Mathematically, the SR task can be formulated as an ill-posed inverse problem: for a given vectorised LR image $\mathbf{g} \in \mathbb{R}^M$, we look for its HR version $\mathbf{u} \in \mathbb{R}^N$ defined on a space of dimension $N = L^2 M$ with magnification factor $L > 1$ which satisfies the following linear degradation model:

$$\mathbf{g} = \mathbf{SHu} + \boldsymbol{\eta}. \tag{1}$$

Here, $\mathbf{S} \in \mathbb{R}^{M \times N}$ stands for the down-sampling operator, $\mathbf{H} \in \mathbb{R}^{N \times N}$ describes blur degradation and $\boldsymbol{\eta}$ denotes the realisation of an Additive White Gaussian Noise (AWGN) random vector with zero mean and covariance matrix $\sigma_\eta^2 \mathbf{I}$.

Due to the ill-posedness of the operator $\mathbf{SH}$, a standard approach for solving (1) consists in encoding prior knowledge about the solution $\mathbf{u}$ and on the data statistics via an energy minimisation approach, so that an approximated solution $\mathbf{u}^* \in \mathbb{R}^N$ is computed by solving

$$\mathbf{u}^* \in \arg\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{SHu} - \mathbf{g}\|_2^2 + \mu R(\mathbf{u}), \tag{2}$$

where the quadratic fidelity term models the presence of AWGN, while the (possibly non-convex) regularisation term $R : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ encodes prior information on the target image $\mathbf{u}$, thus ensuring the stability of the inversion process. The two terms are balanced by a regularisation parameter $\mu > 0$. We refer the reader to [7], [8] and to the references therein for a review on variational approaches for SR problems.

In this work, we choose R so as to promote gradient sparsity, which is often desirable for subsequent image segmentation analyses for which a simplified, edge-preserved version of the original data $\mathbf{g}$ is required. In recent years, sparse, non-convex gradient-based regularisation approaches have become very popular in the context of image reconstruction due to their better ability of preserving sharp edges even in low-contrast scenarios. A significant contribution has been made by Storath et al. in a series of papers [1], [9], [10] where sparsity on the image gradient $\mathbf{Du} \in \mathbb{R}^{2N}$ is promoted by means of ℓ^0 regularisation which reads

$$\|\mathbf{Du}\|_0 := \# \{(\mathbf{Du})_i, i = 1, \dots 2N : (\mathbf{Du})_i \neq 0\}. \tag{3}$$

The use of an ℓ_0 gradient-smoothing has been shown to favour a significant image smoothing which preserves salient image edges and eliminate insignificant details on several imaging problems such as deconvolution, sparse recovery, joint reconstruction and segmentation, image cartoonisation and many more, see, e.g., [11]. In this work, we show that this feature can further be beneficial for classification and labelling analysis upon a suitable super-resolution processing step. To do so, we propose a novel effective numerical scheme solving (1) endowed with convergence guarantees and validate it on several imaging examples with high blur and noise degradation.

P. Cascarano is with the Department of Mathematics, University of Bologna, 40126, Bologna, Italy. Correspondence e-mail: pasquale.cascarano2@unibo.it

L. Calatroni is with CNRS, Université Côte d'Azur, INRIA, I3S, UMR 7271, 06903, Sophia-Antipolis, France.

E. Loli Piccolomini is with the Department of Computer Science and Engineering, University of Bologna, 40126, Bologna, Italy.

A. *Related work*

The vast majority of sparse optimisation approaches for SR problems enforces sparsity either on the signal itself [12] or its representation w.r.t. some basis/overcomplete dictionary [13]. These methods and their non-convex extensions have been shown to be very powerful in several applications such as microscopy imaging [14], where signal-sparsity can be assumed. Different approaches are based on the use of least squares, Fourier series and Tikhonov-type gradient regularisations [7], which favour noise removal at the price of creating smoothing and ringing artefacts which are undesirable in many applications such as object detection where images with sharper edges are preferable for better classification. To overcome this drawback, the use of edge-preserving regularisations based on the idea of gradient sparsity, such as Total Variation (TV) [15]–[18], as well as its fractional [19] and ℓ^0 extension [1], [9], [10], has been proposed. Such methods have shown good performances in many applications, although their convexity (in the case of TV) or their challenging numerical realisation (in the case of its non-convex variants) often limit their practical use and precision. Different approaches for solving the SR problem make use of deep architectures encoding prior information on the desired HR solution from a training set of examples [20]–[22]. In particular, in [23] the authors present a Plug-and-Play (PnP) framework [24] which exploits deep convolutional neural network denoisers embedded in standard optimisation algorithms, such as Alternating Direction of Multipliers (ADMM) or Half-Quadratic Splitting (HQS). Differently from model-based variational approaches, deep learning-based methods do not require an explicit expression of the regularisation term R, since this can be learned directly from the data and adapted to the particular application considered. Those methods have currently reached state-of-the-art performances in many image reconstruction problems, although their theoretical foundation and their stability to noise perturbations still limits their practical use in the case of highly-degraded imaging data.

B. *Contribution*

We consider a variational model for solving (2) where an ℓ^0-gradient regularisation term is considered both in a coupled (isotropic) and decoupled (anisotropic) form, the latter being better suited for directionally-biased images, such as QR scans. To solve the model efficiently, we propose an ADMM algorithm which decomposes the original problem into substeps cheaply solved by means of direct hard-thresholding and standard iterative Conjugate Gradient (CG) linear solvers. Our variable splitting differs from the one introduced by Storath et al. in [1], [9], [10], where the non-convex substeps are solved by means either of approximate graph-cuts approaches [25] or dynamic programming algorithms. For the proposed ADMM algorithm, fixed-point convergence is proved. Up to our knowledge, the same variable splitting has been used only in the case of convex regularisation functions, such as TV, in [17], [18] where convergence to the global minimum is proved.

The proposed SR model is tested on real-world applications (QR scanning, cell detection, land-cover labelling, image car-

toonisation) where a simplified HR version of the given LR image $\mathbf{g}$ is required in view of further analysis, showing that the proposed model improves significantly segmentation and labelling precision in comparison to competitive model- and data-driven approaches.

C. *Organisation of the paper*

In Section II we provide a review of gradient-sparse variational methods for single-image SR. In Section III we present a novel converging ADMM scheme for solving the proposed model along with details on its practical realisation. In Section IV we report some numerical tests on model parameter sensitivity performed on synthetic data. Finally, in Section V we apply our model to some real-world applications such as QR scanning, cell detection, land-cover classification and detail-preserving image cartoonisation. We report the convergence proofs of the proposed ADMM schemes in Appendix A to improve the flow of the manuscript.

II. ℓ^0 GRADIENT-BASED SUPER-RESOLUTION

The use of convex gradient-based regularisations for SR problems dates back to [15], [16], where TV regularisation[1]

$$\|\mathbf{Du}\|_{1,p} = \sum_{i=1}^{N} \left(\|(\mathbf{D_h u})_i\|^p + \|(\mathbf{D_v u})_i\|^p\right)^{1/p}, \quad (4)$$

was employed to promote sparsity on the image gradient $\mathbf{Du} = (\mathbf{D_h u}, \mathbf{D_v u}) \in \mathbb{R}^{N \times 2}$. Note, that for $p \in \{1, 2\}$ anisotropic/isotropic regularisation is promoted, respectively. We remark that fractional generalisations to exponents $1 < p < 2$ are also possible [19].

Gradient-sparsity can be enforced more severely by means of non-convex ℓ^0 gradient smoothing, see, e.g., [11] and [1]. Using an analogous notation as in (4), for $p \in \{1, 2\}$ we thus consider the functional defined by:

$$R(\mathbf{u}) = \|\mathbf{Du}\|_{0,p} \quad (5)$$
$$:= \sum_{i=1}^{N} \begin{cases} \left|(\mathbf{D_h u})_i\right|_0 + \left|(\mathbf{D_v u})_i\right|_0 & \text{for } p = 1, \\ \left|\|(\mathbf{D_h u})_i, (\mathbf{D_v u})_i\|\right|_0 & \text{for } p = 2, \end{cases}$$

where by $|\cdot|_0$ we denote the function:

$$|z|_0 := \begin{cases} 0 & z = 0 \\ 1 & z \neq 0. \end{cases}$$

The functional (5) counts the number of *jumps* of $\mathbf{u}$ in terms of the non-zero values of its gradient magnitude. In particular, in the case $p = 1$ the regulariser independently counts the jumps along the two horizontal and vertical Cartesian directions, whereas for $p = 2$ the gradient magnitudes are taken into account jointly. In both cases, the term $\|\mathbf{Du}\|_{0,p}$ penalises low-amplitude structures while preserving edges in the images, thus favouring sharp piece-wise constant reconstructions which are particularly desirable for image segmentation problems. Notice that $0 \le \|\mathbf{Du}\|_{0,p} \le 2N$ for $p \in \{1, 2\}$.

In the following, we will refer to (5) with $p = 1$ as the *anisotropic* ℓ^0-gradient regularisation model (A-TV^0), while for $p = 2$ we will refer to *isotropic* ℓ^0-gradient regularisation model (I-TV^0).

---

[1]By $\|\cdot\|$ we denote the standard Euclidean modulus.

III. AN EFFICIENT ADMM SPLITTING

For $p \in \{1, 2\}$, we consider the non-smooth and non-convex SR model (2) with the choice (5), that is:

$$\mathbf{u}^* \in \arg\min_{\mathbf{u} \in \mathbb{R}^N} \left\{ \Phi(\mathbf{u}; \mu, p) := \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \mu\|\mathbf{Du}\|_{0,p} \right\}. \tag{6}$$

Existence of solutions for (6) is guaranteed by the following theorem whose proof can be found in [1, Theorem 1] for a general forward operator $\mathbf{A}$.

Theorem 1: The solution set of both the anisotropic ($p = 1$) and isotropic ($p = 2$) problem (6) is non-empty.

To solve numerically problem (6) we propose a novel iterative *alternating direction method of multipliers* (ADMM) based on a suitable variable splitting. We separate the description for the anisotropic and isotropic case. For both cases and upon suitable conditions, fixed-point convergence of the ADMM iterates is proved (Theorems 2 and 3).

A. ADMM for A-TV^0

For $p = 1$, we can rewrite the unconstrained minimisation problem (6) in the following equivalent constrained form:

$$\arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \mu(\|\mathbf{t}\|_0 + \|\mathbf{s}\|_0)$$
$$s.t. \quad \mathbf{t} := \mathbf{D}_h \mathbf{u}, \quad \mathbf{s} := \mathbf{D}_v \mathbf{u}$$

where $\mathbf{t}, \mathbf{s} \in \mathbb{R}^N$ represent the horizontal/vertical gradient components, respectively. We then define the augmented Lagrangian function:

$$L_{\beta_t, \beta_s}(\mathbf{u}; \mathbf{t}, \mathbf{s}, \boldsymbol{\lambda}_t, \boldsymbol{\lambda}_s) := \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \mu\|\mathbf{t}\|_0 + \mu\|\mathbf{s}\|_0$$
$$+ \langle \boldsymbol{\lambda}_t, \mathbf{D}_h \mathbf{u} - \mathbf{t} \rangle + \frac{\beta_t}{2}\|\mathbf{D}_h \mathbf{u} - \mathbf{t}\|_2^2 + \langle \boldsymbol{\lambda}_s, \mathbf{D}_v \mathbf{u} - \mathbf{s} \rangle$$
$$+ \frac{\beta_s}{2}\|\mathbf{D}_v \mathbf{u} - \mathbf{s}\|_2^2 \tag{7}$$

where β_t and β_s are two positive penalty parameters and $\boldsymbol{\lambda}_t$ and $\boldsymbol{\lambda}_s$ are the vectors of Lagrange multipliers related to the auxiliary variables $\mathbf{t}$ and $\mathbf{s}$, respectively. By letting the two parameters β_t, β_s increase along the iterations (we will provide specific growth conditions in the following Theorem 2), we can then minimise (7) w.r.t. $\mathbf{t}, \mathbf{s}$ and $\mathbf{u}$ by iterating the following scheme:

$$\begin{cases} \mathbf{t}^{k+1} \in \arg\min_{\mathbf{t}} \mu\|\mathbf{t}\|_0 + \frac{\beta_t^k}{2}\|\mathbf{t} - (\mathbf{D}_h \mathbf{u}^k + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})\|_2^2 & (8) \\[2mm] \mathbf{s}^{k+1} \in \arg\min_{\mathbf{s}} \mu\|\mathbf{s}\|_0 + \frac{\beta_s^k}{2}\|\mathbf{s} - (\mathbf{D}_v \mathbf{u}^k + \frac{\boldsymbol{\lambda}_s^k}{\beta_s^k})\|_2^2 & (9) \\[2mm] \mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \\[1mm] \quad + \frac{\beta_t^k}{2}\|\mathbf{D}_h \mathbf{u} - (\mathbf{t}^{k+1} - \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})\|_2^2 + \frac{\beta_s^k}{2}\|\mathbf{D}_v \mathbf{u} - (\mathbf{s}^{k+1} - \frac{\boldsymbol{\lambda}_s^k}{\beta_s^k})\|_2^2 & (10) \\[2mm] \boldsymbol{\lambda}_t^{k+1} = \boldsymbol{\lambda}_t^k - \beta_t^k(\mathbf{t}^{k+1} - \mathbf{D}_h \mathbf{u}^{k+1}) & (11) \\[2mm] \boldsymbol{\lambda}_s^{k+1} = \boldsymbol{\lambda}_s^k - \beta_s^k(\mathbf{s}^{k+1} - \mathbf{D}_v \mathbf{u}^{k+1}), & (12) \end{cases}$$

where a gradient ascent update of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\lambda}_s$ is also applied.

Under suitable growth assumptions, the sequences (8), (9), (10) converge to a fixed point (see Appendix A for the proof).

Theorem 2: Let the ADMM iterations (8)-(12) be defined under the following conditions:

A.1 $(\beta_t^k), (\beta_s^k)$ are increasing sequences such that $\sum_{k=1}^{+\infty} \sqrt{\frac{k}{\beta_t^k}} < +\infty$, $\sum_{k=1}^{+\infty} \sqrt{\frac{k}{\beta_s^k}} < +\infty$ and $\frac{\beta_s^k}{\beta_t^k} \to c \neq 0$.

A.2 $\mathbf{D}_h$ and $\mathbf{D}_v$ are full rank.

Then, the sequences $(\mathbf{t}^k), (\mathbf{s}^k), (\mathbf{u}^k)$ converge, i.e.:

$$\mathbf{t}^k \longrightarrow \mathbf{t}^*, \quad \mathbf{s}^k \longrightarrow \mathbf{s}^*, \quad \mathbf{u}^k \longrightarrow \mathbf{u}^*,$$

with $\mathbf{t}^* = \mathbf{D}_h \mathbf{u}^*$ and $\mathbf{s}^* = \mathbf{D}_v \mathbf{u}^*$.

We remark that the full rank assumption on the operators $\mathbf{D}_h$ and $\mathbf{D}_v$ is verified, for instance, if Dirichlet boundary conditions are assumed. A sufficient condition which guarantees the required growth of the penalty sequences is $\beta_t^k = \beta_s^k = O(k(1 + \epsilon)^k)$, $0 < \epsilon \ll 1$.

B. ADMM for for I-TV^0

For $p = 2$ we can write problem (6) in the following equivalent constrained form:

$$\arg\min_u \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \mu\sum_{i=1}^N \big|\|\mathbf{z}_i\|\big|_0 \tag{13}$$
$$s.t. \quad \mathbf{z} := \mathbf{Du}$$

where $\mathbf{z}_i := \big((\mathbf{D}_h \mathbf{u})_i, (\mathbf{D}_v \mathbf{u})_i\big) \in \mathbb{R}^2$, for each $i = 1, \ldots, N$. The augmented Lagrangian function reads in this case:

$$L_\beta(\mathbf{u}; \mathbf{z}, \boldsymbol{\lambda}) := \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \mu\sum_{i=1}^N \big|\|\mathbf{z}_i\|\big|_0$$
$$+ \langle \boldsymbol{\lambda}, \mathbf{Du} - \mathbf{z} \rangle + \frac{\beta}{2}\|\mathbf{Du} - \mathbf{z}\|_2^2 \tag{14}$$

where $\beta > 0$ is a scalar penalty parameter and $\boldsymbol{\lambda} \in \mathbb{R}^{2 \times N}$ is the Lagrange multiplier vector. As above, by letting the penalty parameter increases along the iterations at a certain growth (see the following Theorem 3), we seek for minimisers of (13) by iterating the following scheme:

$$\begin{cases} \mathbf{z}^{k+1} \in \arg\min_{\mathbf{z}} \mu\sum_{i=1}^N \big|\|\mathbf{z}_i\|\big|_0 + \frac{\beta^k}{2}\|\mathbf{z} - (\mathbf{Du}^k + \frac{\boldsymbol{\lambda}^k}{\beta^k})\|_2^2 & (15) \\[2mm] \mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \frac{\beta^k}{2}\|\mathbf{Du} - (\mathbf{z}^{k+1} - \frac{\boldsymbol{\lambda}^k}{\beta^k})\|_2^2 & (16) \\[2mm] \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta^k(\mathbf{z}^{k+1} - \mathbf{Du}^{k+1}). & (17) \end{cases}$$

For this scheme, a similar result as the one in Theorem 2 holds (see Appendix A for a sketch of the proof).

Theorem 3: Let the ADMM iterations (15)-(16) be defined under the following conditions:

I.1 (β^k) is an increasing sequence such that $\sum_k^{+\infty} \sqrt{\frac{k}{\beta^k}} < +\infty$

I.2 $\mathbf{D}$ is full rank.

Then, $(\mathbf{z}^k) \longrightarrow \mathbf{z}^*$, $(\mathbf{u}^k) \longrightarrow \mathbf{u}^*$ and $\mathbf{z}^* = \mathbf{Du}^*$.

We remark that in order to guarantee the full rank of the operators $\mathbf{D}_h$ and $\mathbf{D}_v$, an artificial image padding of the image can be considered. Furthermore, our numerical experiments, however, showed numerical convergence even when periodic boundary conditions are used. A theoretical convergence proof

in this case is left for future research. As far as the growth condition on the penalty parameters is concerned, we remark that in [1] a geometric growth was assumed. Unfortunately, this is not enough for our theoretical convergence result to hold, as oscillations may appear. We comment more on this in Section V-A.

C. Efficient solution of the ADMM subproblems

We report here some practical details on the the efficient solutions of the subproblems (8)-(10) and (15)-(16).

Solution of ℓ^0 subproblems: Due to decomposability of the ℓ^0 term, solving problems (8),(9) corresponds to solve the N one-dimensional $\ell^2 - \ell^0$ problems

$$\arg\min_{\mathbf{t}_i \in \mathbb{R}} \delta|\mathbf{t}_i|_0 + (\mathbf{t}_i - \mathbf{f}_i)_2^2 \qquad (18)$$

where $\delta = \frac{2\mu}{\beta_t^k}$ and $\mathbf{f}_i = (\mathbf{D}_h \mathbf{u}^k + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})_i$ for (8), while $\delta = \frac{2\mu}{\beta_s^k}$, $\mathbf{f}_i = (\mathbf{D}_v \mathbf{u}^k + \frac{\boldsymbol{\lambda}_s^k}{\beta_s^k})_i$ for (9). As far as the problem (15) is concerned, it similarly reduces to the solution of the N two-dimensional ℓ^0-regularised problems

$$\arg\min_{\mathbf{z}_i \in \mathbb{R}^2} \delta\big\|\|\mathbf{z}_i\|\big\|_0 + \|\mathbf{z}_i - \mathbf{f}_i\|_2^2 \qquad (19)$$

where $\delta = \frac{2\mu}{\beta^k}$ and $\mathbf{f}_i = (\mathbf{D}_h \mathbf{u}_i^k + \frac{(\boldsymbol{\lambda}^k)_{1,i}}{\beta^k}, \mathbf{D}_v \mathbf{u}_i^k + \frac{(\boldsymbol{\lambda}^k)_{2,i}}{\beta^k})$. Solving (18) and (19) corresponds to compute the proximal mapping of $|\cdot|_0$ with parameter δ evaluated in $\mathbf{f}_i$, which is nothing but the 1D [26] and 2D [11] hard-thresholding operator, respectively.

Solution of the quadratic subproblems: The first order optimality conditions of problems (10) and (16) lead to the solution of large-size linear systems, whose coefficient matrix is symmetric and positive definite. To solve them efficiently, we make use of Conjugate Gradient (CG) algorithm with a warm-start initialisation at every iteration. We remark that, due to the presence of the downsampling operator $\mathbf{S}$, the use of more efficient solvers based, for instance, on discrete Fourier transforms are here not possible, as the product matrix $\mathbf{SH}$ does not have a block-circulant structure. However, under suitable assumptions on the down-sampling operator $\mathbf{S}$, the problem admits a closed form solution [27].

D. Comparisons with previous splittings

The variable splitting and the ADMM iterations considered above are different from the ones considered in [1], [9], [10] where the choice $\mathbf{z} = \mathbf{u}$ in (13) is made. Our choice avoids the presence of the gradient operator in the ℓ^0-based problems (8)-(9) and (15), leading to the faster computation of their solution by direct solvers without requiring the use of approximate solvers based on approximate graph-cut algorithms [1]. These latter algorithms have well-known drawbacks such as strong dependence on the initialisation and require an approximate inner solver [9], [28]. As an alternative, in [9], the isotropic substep is solved by a set of anisotropic problems along the diagonal or knight-move directions, each of which is computed by dynamic programming algorithms with computational cost $O(N^2)$ compared to $O(N)$ in our approach.

IV. Implementation notes

1) Operators: For the following examples, we simulate the LR data from ground-truth HR images by applying the forward model (1) where the action of the blur matrix $\mathbf{H}$ is computed by assuming a Gaussian PSF with zero mean and standard deviation σ_H which will be specified later on. As $\mathbf{S}$, we consider the discretised 2D Lanczos down-sampling operator [29] inbuilt in the MATLAB function `imresize`. Finally, we consider AWGN with zero mean and standard deviation $\sigma_{\boldsymbol{\eta}}$ whose values will be made precise in the following.

2) Comparisons: We compare our results with the ones obtained by gradient-sparse regularisation models such as convex isotropic TV (I-TV) [16], non-convex capped TV (c-TV) [30] and anisotropic fractional TV [31] which, for consistency, have been implemented within the same ADMM optimisation framework. We also provide comparisons with standard bicubic interpolation and to the SR approach based on sparse representation (SrSR) proposed in [13], upon a suitable training of the dictionaries. Finally, we add comparisons with the results obtained by two state-of-the-art Deep Learning-based approaches. The former is the Content Adaptive Re-sampler (CAR) [22] convolutional neural network, which is characterised by a downsampler-upsampler structure. For that, we use a pre-trained model [2] taking into account only the trained upsampler part. The latter is the Image Restoration Convolutional Neural Network (IRCNN) [23], which is a Plug and Play (PnP) method based on HQS optimisation.

3) Initialisation, parameters and evaluation metrics: We initialise $\mathbf{u}^0$ in our model as $\mathbf{u}^0 = \mathbf{S}^T \mathbf{g}$. Given the non-convexity of problem (6), the choice of a wise initialisation is important. We tested several ones (the aforementioned one, the zero image and the I-TV initialisation) and kept the one providing the best results. The variables $\mathbf{t}^0, \mathbf{s}^0, \mathbf{z}^0$ as well as $\boldsymbol{\lambda}_t^0, \boldsymbol{\lambda}_s^0, \boldsymbol{\lambda}^0$ in (8)-(12) and in (15)-(17) were set to $\mathbf{0}$. To ensure the convergence results provided by Theorems 2 and 3, the penalty sequences are chosen as $(\beta^k) = k(1+\epsilon)^k$ with $\epsilon = 10^{-4}$. Note that for such small choice of ϵ, $k(1+\epsilon)^k \approx k$, i.e. the growth of (β^k) is almost linear. The process is stopped when the relative change between consecutive iterates $\mathbf{u}^k$ is lower than 10^{-3}.

For simulated data, we evaluate the quality of the SR outputs by means of Peak-Signal-to-Noise-Ratio (PSNR) and Structure Similarity index (SSIM) as well as the Jaccard index (Jac), an evaluation metric in the range $[0, 1]$ measuring the ratio between correctly detected points and false detections. We remark that choosing the right evaluation metric for SR problems is not trivial, see, e.g., [32] for a review. While PSNR and SSIM are good choices to quantify reconstruction quality, the Jaccard index is more appropriate for segmentation purposes as it assesses correct versus false pixel localisation.

V. Numerical Experiments

We report here several experiments performed on synthetic and real data. All the experiments are executed on a PC Intel(R) Core(TM) i5-6200U CPU @ 2.30 GHz 2.40GHz with

---

[2]https://github.com/sunwj/CAR

8.00Gb RAM using Matlab R2018b and Python 3. The codes are available at https://github.com/pcascarano/PottsSR.

A. Computational analysis on synthetic data

We first validate numerically the convergence properties of the proposed ADMM algorithms and comment on their parameter sensitivity.

For this first example, LR data were generated by applying (1) to the HR 428×600 grayscale image in Figure 1 (a). Namely, Gaussian blur with $\sigma_H = 1$, $L = 4$ down-sampling and AGWN with standard deviation $\sigma_\eta = 0.01$ were applied to get the LR image in Figure 1 (b). In Figure 1 (c)-(f) we report the results computed by the anisotropic (A-TV^0) and isotropic (I-TV^0) ℓ^0-gradient model for two different values of the regularisation parameter $\mu \in \{0.005, 0.01\}$. As expected, the jump-sparse regularisation flattens out many details in the reconstruction, promoting a cartoon-like reconstruction: the higher the regularisation parameter μ, the more simplified the reconstruction. We further add a close-up of two ROIs: the blue square contains both fine details (filaments, yellow arrows) and corner points (green arrows), the red one textured details. The directional bias of the A-TV^0 regularisation along the horizontal and vertical direction is here clearly visible. We report in the captions of Figure 1 (c)-(f) the values $\|\mathbf{Du}^*\|_{0,1}$ and $\|\mathbf{Du}^*\|_{0,2}$ which correspond to the number of gradient jumps on the output image. Clearly, choosing a larger μ promotes more jump-sparsity, so that the number of jumps of $\mathbf{u}^*$ is smaller.

We then validate the algorithmic convergence behaviour w.r.t. the choice of the penalty sequences $(\beta_t^k), (\beta_s^k), (\beta^k)$. Namely, in Figure 2 (a) and 2 (b) we report the behaviour of the objective functions $\Phi(\mathbf{u}^k; \mu, p)$ in (6) along the ADMM iterations for different choices of the penalty sequences (left). For both cases $p = 1$ and $p = 2$ we choose $\beta^k = \beta_t^k = \beta_s^k \equiv 10$ for all k (blue line), $\beta^k = \beta_t^k = \beta_s^k = k^{0.5}$ (red line) and $\beta^k = \beta_t^k = \beta_s^k = k(1 + \epsilon)^k$ with $\epsilon = 10^{-4}$ (yellow line). On the same plots we further show the decay of the quadratic data term (right). We observe that when the penalty sequence fulfil the required growth condition, then the convergence is nicely monotone, whereas for the other two choices, the decay exhibits oscillations while preserving a globally decreasing trend. Numerically, this suggests that possibly less severe growth conditions may be employed, such as a sufficiently large constant values of the penalty parameters. A further study on this is left for future research.

To confirm the improved computational performance of our ADMM algorithm w.r.t. to the one proposed in [9] and adapted to solve the SR problem (6), we report in Table I a comparison table both in terms of number of iterations-to-convergence and computational times. We stress that the poor performance of the ADMM algorithm in [9] is due here to the large computational cost required to solve the ℓ^0 gradient steps via inner optimisation routines. This, combined with the use of CG solvers (required for the SR problem under consideration as no Fourier-based approaches can be used in general) makes the overall cost much higher in comparison to our more explicit splitting.

Table I: Iterations till convergence (`iter`) and computational times (in seconds) for different methods solving (6).

Method	[9]	A-TV^0	I-TV^0
`iter`	1905	63	59
time (s)	2866.31	214.83	195.99

(a) HR

(b) LR (x4)

(c) $\|\mathbf{Du}^*\|_{0,1}$=26822, $\mu = 0.005$

(d) $\|\mathbf{Du}^*\|_{0,2}$=24067, $\mu = 0.005$

(e) $\|\mathbf{Du}^*\|_{0,1}$=19059, $\mu = 0.01$

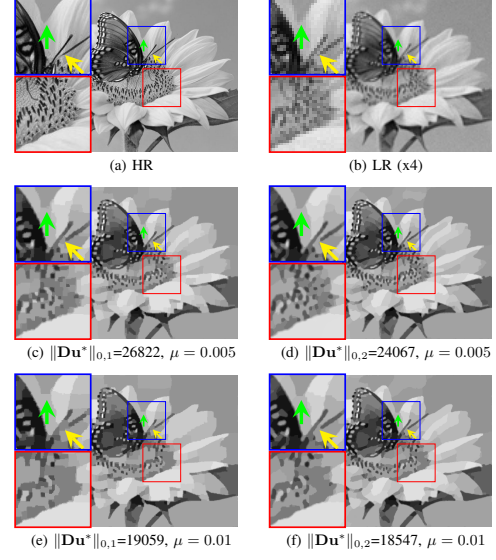(f) $\|\mathbf{Du}^*\|_{0,2}$=18547, $\mu = 0.01$

Figure 1: Results obtained for $\mu \in \{0.005, 0.01\}$ by A-TV^0 and I-TV^0 on a synthetic image.
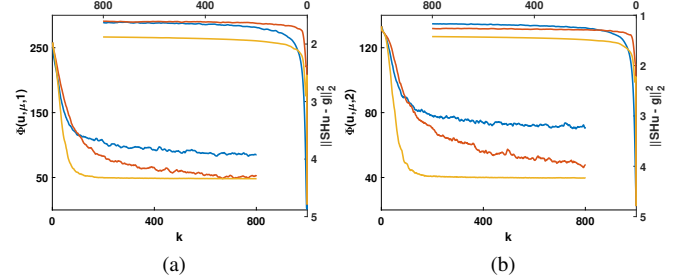

(a)

(b)

Figure 2: Values of the cost function in (6) (left y-axis) and of the fidelity term (right y-axis) along iterations in the two cases $\Phi(\mathbf{u}^k; \mu, 1)$ (a) and $\Phi(\mathbf{u}^k; \mu, 2)$ (b), for $\mu = 0.01$. The penalty sequences are chosen as $\beta^k = \beta_t^k = \beta_s^k \equiv 10$ (blue), $\beta^k = \beta_t^k = \beta_s^k = k^{0.5}$ (red), $\beta^k = \beta_t^k = \beta_s^k = k(1 + \epsilon)^k$ with $\epsilon = 10^{-4}$ (yellow).

B. Real-world applications

We now report the results obtained by applying the proposed model to different real-world applications where a SR version of the given LR image is required for further image analysis.

1) QR code recognition: The first application we consider is the problem of QR detection. As described, e.g., in [4], images of QR codes are often scanned by means of portable devices with limited resolution. Furthermore, QR scans are often taken from a distance and in non-optimal optical conditions so that blur and noise further limit the amount of visible information, thus making the use of artefact-free SR approaches crucial.

For the following tests, we first generate a binary QR code image of size 250×250 by using a freely available QR

code generator [3], then we simulate several LR acquisitions for different levels of degradation. We consider three test cases: $\sigma_\eta = 0.01$ and $\sigma_H = 1$ (TEST 1), $\sigma_\eta = 0.05$ and $\sigma_H = 1$ (TEST 2), and $\sigma_\eta = 0.01$ and $\sigma_H = 4$ (TEST 3). We compare the results obtained by our model with the ones obtained by the models recalled in Section IV-2. For each method, we select the model parameters maximising the Jaccard index. To avoid non-binary outputs (required for Jaccard index computations), for all models we post-process the SR results by means of an adaptive Otsu thresholding and re-compute the evaluation metrics on the binarised output, see Table II.

In Figure 3 we report the results obtained by the different methods for the TEST 2 image before (red frame) and after (blue frame) binarisation. We observe that due to the sharp nature of the the TV^0 regulariser, the A-TV^0 and I-TV^0 results are almost binary so they do not benefit much from the post-processing step in terms of Jaccard index values as much as the other methods do. In Figure 4 we report a zoom of the best results obtained before binarisation by all methods starting from the highly corrupted TEST 3 LR image.
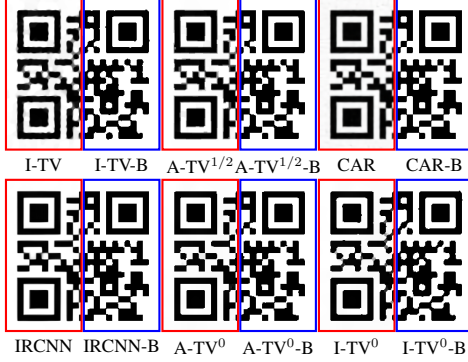

Figure 3: QR SR results obtained by different methods on TEST2 image before (red frames) and after (blue frames) binarisation.
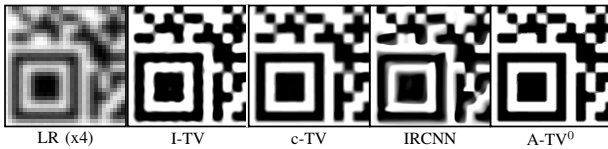

Figure 4: Details of QR code SR outputs for TEST 3 image.

The quantitative evaluation of the results in terms of PSNR, SSIM and Jaccard index for the three different test cases is reported in Table II. Without any binarisation, the A-TV^0 model outperforms all the others as far as the PSNR, SSIM and Jaccard indices are concerned. The simplified geometry of the QR images considered (i.e. the sole presence of horizontal/vertical edges) makes in fact this kind of data tailored for such geometrically-biased regularisations. Furthermore, the highly non-convex jump-sparsification forces the ouptut to be almost binary, without the need of any further post-processing binarisation, as it is required by all the other regularisations to achieve comparable (if not better) quality scores. This simple example shows that the image simplification intrinsically

favoured by the use of TV^0 regularisers shall limits the need of post-processing techniques in view of further segmentation analysis. Concerning the comparison with the SrSR method, we remark that the dictionaries used have been trained on a dataset of 40 QR codes. Good results are obtained for this model only after binarisation. As far as the deep-learning results are concerned, we remark that the CAR network in this experiment is used in a transfer learning mode, with no noisy nor blurred images observed in the training phase. For a fairer comparison, we thus consider the IRCNN PnP network which is capable to handle different levels of degradations, although it is shown to fail in the presence of highly-degraded data, see Figure 4.

Table II: Quantitative evaluation of SR models performance on QR for three different TEST images and methods. By "-B" we denote results after binarisation. In each column we colour red the best method, blue the second-best.

LR	Method	PSNR	PSNR-B	SSIM	SSIM-B	Jac
TEST 1	I-TV^0	22.5199	29.0809	0.9423	0.9873	0.9980
	A-TV^0	32.5943	35.8478	0.9913	0.9989	0.9999
	I-TV	23.3845	26.3357	0.9489	0.9762	0.9963
	c-TV	19.4522	36.7496	0.8849	0.9977	0.9997
	A-$TV^{1/2}$	18.6328	36.7496	0.8594	0.9989	0.9997
	CAR	20.2460	27.8163	0.8159	0.9801	0.9966
	IRCNN	25.0589	35.3363	0.9622	0.9992	0.9995
	SrSR	17.8837	26.4318	0.7970	0.9779	0.9964
	Bicubic	14.5489	19.9651	0.5974	0.9211	0.9838
TEST 2	I-TV^0	19.3318	18.6308	0.8766	0.9156	0.9781
	A-TV^0	22.6887	22.6256	0.9242	0.9653	0.9912
	I-TV	18.1101	18.9848	0.8012	0.9171	0.9798
	c-TV	18.7331	21.3473	0.8211	0.9595	0.9882
	A-$TV^{1/2}$	19.2182	22.5108	0.8664	0.9660	0.9910
	CAR	18.1320	26.7831	0.7493	0.9805	0.9906
	IRCNN	21.4314	26.3968	0.9057	0.9850	0.9902
	SrSR	17.4148	21.4698	0.6891	0.9465	0.9886
	Bicubic	14.4037	17.8804	0.5346	0.8806	0.9739
TEST 3	I-TV^0	18.3763	19.7532	0.8634	0.9294	0.9831
	A-TV^0	19.2908	21.9341	0.8861	0.9556	0.9897
	I-TV	17.9552	20.1585	0.8222	0.9282	0.9846
	c-TV	16.9580	22.4648	0.7915	0.9605	0.9917
	A-$TV^{1/2}$	17.0785	20.6874	0.7706	0.9372	0.9863
	CAR	11.1809	11.5412	0.4057	0.6342	0.8887
	IRCNN	14.2915	12.5640	0.6342	0.6565	0.9133
	SrSR	14.5796	19.1445	0.5483	0.9093	0.9806
	Bicubic	10.8695	10.1082	0.3838	0.5194	0.8450

2) Cell detection: Standard light-microscopes suffer from a limited resolving power which often causes blur artefacts and limits spatial resolution in images. In such conditions, the performance of simple segmentation algorithms extracting isolated cells as well as cell clusters is often very limited and may benefit significantly from the use of a joint super-resolution image restoration pre-processing. We thus test the proposed ℓ^0-gradient SR model to segment a dataset of 30 light-microscope images extracted from the EVICAN dataset [33]. We apply the I-TV^0 model and its competitors on LR acquisitions obtained by applying the linear degradation model (1) to the original images, considered here as Ground Truth (GT) with the following values of parameters: $L = 3$, $\sigma_H = 2$ and $\sigma_\eta = 0.02$. In Figure 5 we show the results for one test image in the dataset. Due to our interest in analysing the effectiveness of the proposed methods in pre-processing images for segmentation, we compute for each SR output image a binary mask by applying the cell-segmentation

(a) GT   (b) LR(x3)   (c) A-TV$^{1/2}$ - Jac=0.90

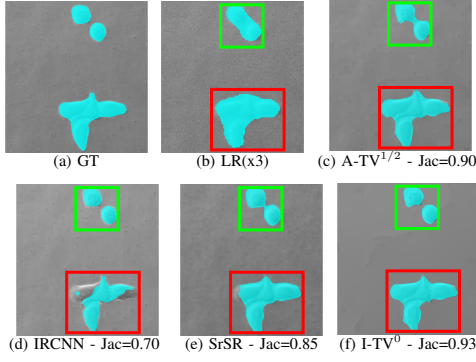(d) IRCNN - Jac=0.70   (e) SrSR - Jac=0.85   (f) I-TV^0 - Jac=0.93

Figure 5: Cell detection results. In (b)-(f) the green and red squares indicate two isolate cells and a cell cluster, respectively. Computed masks are coloured cyan. The Jaccard index (Jac) is reported.

Matlab toolbox [4] based on edge detection and morphology. For the different methods, the segmented regions are shown in Figure 5 (c)-(e), while in Table III the confidence intervals (95% of confidence) of the PSNR, SSIM and Jaccard values computed on the whole dataset are reported. Thanks to its strong smoothing properties, the ℓ^0-gradient sparsity enforced by the I-TV^0 method allows for a better detection of the two isolated cells (green boxes) as well as the cell cluster (red boxes), resulting in higher Jaccard index values, as expected.

Table III: Quantitative comparisons of SR and segmentation performance among different methods on the EVICAN dataset. Confidence intervals with 95% of confidence.

Method	PSNR	SSIM	Jaccard
Bicubic	[30.3383,32.8808]	[0.6841,0.7857]	[0.3594,0.5434]
I-TV	[30.5078,35.6530]	[0.6718,0.8796]	[0.4265,0.7070]
A-TV$^{1/2}$	[32.0277,35.6011]	[0.7473,0.8775]	[0.7767,0.8541]
I-TV^0	[31.9137,35.5359]	[0.7458,0.8778]	[0.7775,0.8622]
SrSR	[30.9591,35.8848]	[0.6711,0.8855]	[0.6594,0.8102]
IRCNN	[32.9723,35.9698]	[0.7871,0.9070]	[0.5501,0.7531]

3) Land-cover classification: Multi-Spectral Imaging (MSI) is fundamental in the field of land-cover mapping and classification [34] thanks to its ability of quantify different types of information about the objects in the recorded scene, such as their physical composition and their temperature. Existing MSI segmentation techniques exploit these properties to label each pixel of the given image within a class, thus producing a final 2D labelled image. These maps are essential in many sustainability-related applications and monitoring purposes for detecting land-cover changes (e.g. deforestation) over the years at the same geographical location, which cannot be done directly by simply looking at the raw MSI data (see [35] and references therein). Among the many existing open-source MSI datasets, we consider here the National Agriculture Imagery Program (NAIP) [36] dataset and the Hamlin Beach State Park (HBSP) [37] dataset and apply SR methods to increase the spatial resolution of the given MSI data so as to produce an output image which could be easily segmented. The need of a SR model in this specific application is justified by the physical limitations preventing
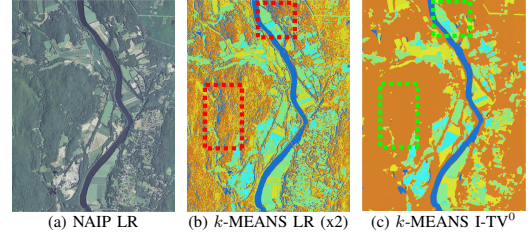

(a) NAIP LR   (b) k-MEANS LR (x2)   (c) k-MEANS I-TV^0

Figure 6: k-means segmentation ($k = 5$) of MSI data. (a) LR image (x2) (b) k-means classification of LR image (c) k-means classification of I-TV^0 output. The red/green boxes show possible misclassifications.

HR acquisitions, such as the limited spatial resolution in the infrared band [38]. Moreover, as shown in the previous examples, a SR I-TV^0 pre-processing step is expected to improve image classification results, which can be performed several algorithms. Here, we consider standard k-Means segmentation and the state-of-the-art U-Net neural network [39], showing that in both cases the use of a SR I-TV^0 significantly improves classification accuracy.

In the first experiment we consider a LR image [5] from the NAIP dataset (Figure 6 (a)). We first run the k-Means algorithm directly on this image, choosing empirically the number of classes to be $k = 5$. The classification obtained looks speckled and significant classification errors occur (see 6 (b)). In Figures 6 (c)-(d), the classification result obtained by applying k-Means to I-TV^0 SR reconstructions (with $L = 2$) is reported. Note that the classification performed on the SR image appears much more reliable, with reduced classification errors (red boxes).

As a second test, we use the SR I-TV^0 model to pre-process an image from the validation set of the HBSP dataset before applying the unsupervised U-Net segmentation approach for multi-class segmentation. To do so, we simulate a LR MSI acquisition of size $440 \times 350 \times 6$ and apply the SR model (with $L = 2$) to each individual channel. As above, we use the U-Net,which has been trained on a dataset of HR MSIs, both on the given LR MSI and on the computed SR reconstruction, see Figure 7. The quality of the U-Net segmentation is significantly improved when a pre-processing with SR I-TV^0 is performed. When applied to the given LR image, U-Net is indeed not capable to differentiate the group of trees (blue) from the grass (red) as it happens after the application of the SR I-TV^0 smoothing.

4) Detail-preserving image cartoonisation: In [11] ℓ^0-gradient regularisation was extensively shown to be effective on several image smoothing applications such as image cartoonisation and JPG compression artefact removal. Here, we consider a scenario where analogous tasks are performed along with a resolution improvement. To do so, we consider an RGB LR cartoon-type image of size 170×170 with not discernible details due to noise and blur artefacts caused by image compression and apply gradient-sparse SR models. As no ground truth is available for this example, for all models we empirically select the parameters producing the best visual

---

[4] https://www.mathworks.com/help/images/detecting-a-cell

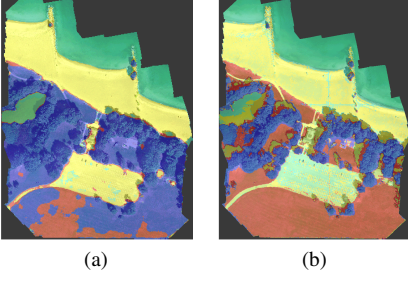[5] Image identification number: M 4207221 NW 18 1 20120709

Figure 7: Results of MSI segmentation by U-NET. (a) Result on the given LR image (x2). (b) Result on the I-TV^0 reconstruction.

output. In Figure 8 we report two close-ups of the computed SR reconstructions marked by blue and red boxes. The blue box highlights small details which are poorly discernible in the LR image, while the red box considers a patch of the face with some blunt edges and a small (but meaningful!) face mole (see green box). We see that both A-TV^0 and I-TV^0 reconstructions are sharper and more cartoonised than the ones obtained by the other models. Furthermore, the well-known I-TV and c-TV loss of contrast reconstruction artefact makes small details hardly discernible. Due to the high-level of compression artefacts, both IRCNN and CAR perform poorly.

VI. CONCLUSIONS

We considered a variational model with ℓ^0 gradient-sparsity-promoting regularisation combined with a quadratic data fidelity for single-image super-resolution of images corrupted by blur and Gaussian noise. The use of non-convex jump-sparse regularisations has been applied in [1] to general 1D inverse problems and subsequently applied in [9], [10] to joint image segmentation and reconstruction problems. To overcome the computational limitations required by the use of ADMM splitting strategies considered in these works, we propose a novel ADMM algorithm allowing for the efficient numerical solution of the models by means of direct hard-thresholding or standard CG solvers. For the proposed schemes, we prove fixed-point convergence results assuming specific growth conditions on the sequence of penalty parameters. We validate our model on synthetic data and test it on real-world examples where gradient-sparse super-resolved outputs are required in view of a subsequent accurate detection/segmentation step (such as QR code recognition [4], cell detection and land-cover classification [34]). By numerous comparisons with convex and non-convex variational approaches, and with state-of-the-art deep learning methods [22], [23], we show that the proposed approach significantly improves classification precision, while limiting at the same times smoothing and loss-of-contrast artefacts in comparison with classical convex regularisations.

Further work should address the use of analogous algorithms for the joint modelling of SR and segmentation problems via, e.g., Mumford-Shah functionals [9]. Furthermore, the extension of the convergence results to other gradient discretisations and to less restrictive growth conditions for the sequence of penalty parameters is envisaged.

APPENDIX A
CONVERGENCE ANALYSIS

We report here a complete convergence proof of Theorem 2 and a sketch of the proof of Theorem 3, which is based on similar arguments.

A. Proof of Theorem 2

Proof: We consider the ADMM sequences $(\mathbf{u}^k), (\mathbf{t}^k), (\mathbf{s}^k)$, defined in (8)-(10). We want to show that there exists $\mathbf{u}^*$ such that:

$$\mathbf{u}^k \to \mathbf{u}^*, \quad \mathbf{t}^k \to \mathbf{D}_h \mathbf{u}^*, \quad \mathbf{s}^k \to \mathbf{D}_v \mathbf{u}^*.$$

To shorten the proof, we remark that everything proved for the sequences $(\mathbf{t}^k), (\beta_t^k), (\boldsymbol{\lambda}_t^k)$ and $(\mathbf{D}_h \mathbf{u}^k)$ can be deduced for $(\mathbf{s}^k), (\beta_s^k), (\boldsymbol{\lambda}_s^k)$ and $(\mathbf{D}_v \mathbf{u}^k)$ in the same way.

We start defining the following functionals:

$$G_k^h(\mathbf{t}) := \mu\|\mathbf{t}\|_0 + \frac{\beta_t^k}{2}\|\mathbf{t} - (\mathbf{D}_h \mathbf{u}^k + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})\|_2^2,$$

$$F_k(\mathbf{u}) := \frac{1}{2}\|\mathbf{SHu} - \mathbf{g}\|_2^2 + \frac{\beta_t^k}{2}\|\mathbf{D}_h \mathbf{u} - (\mathbf{t}^{k+1} - \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})\|_2^2 +$$
$$+ \frac{\beta_s^k}{2}\|\mathbf{D}_v \mathbf{u} - (\mathbf{s}^{k+1} - \frac{\boldsymbol{\lambda}_s^k}{\beta_s^k})\|_2^2.$$

Step 1: There holds:

$$\|\mathbf{t}^{k+1} - \mathbf{D}_h \mathbf{u}^k - \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_2 \leq \sqrt{\frac{2\mu N}{\beta_t^k}}. \tag{20}$$

This inequality can be trivially shown by the minimality of $\mathbf{t}^{k+1}$ in (8) which entails $G_k^h(\mathbf{t}^{k+1}) \leq G_k^h(\mathbf{D}_h \mathbf{u}^k + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})$, therefore we get:

$$\mu\|\mathbf{t}^{k+1}\|_0 + \frac{\beta_t^k}{2}\|\mathbf{t}^{k+1} - (\mathbf{D}_h \mathbf{u}^k + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k})\|_2^2$$
$$\leq \mu\|\mathbf{D}_h \mathbf{u}^k + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_0 \leq \mu N,$$

by definition of $\|\cdot\|_0$, where we recall N is the dimension of the vector $\mathbf{u}^k$. By neglecting the first term on the Left Hand Side (LHS) of the above inequality, we deduce (20).

Step 2: From the minimality of $\mathbf{u}^{k+1}$ in (10) we have: $F_k(\mathbf{u}^{k+1}) \leq F_k(\mathbf{u}^k)$ for every k. By definition of F_k and applying (20) and its analogous related to the sequences $(\mathbf{s}^k)$, $(\beta_s^k), (\boldsymbol{\lambda}_s^k)$ and $(\mathbf{D}_v \mathbf{u}^k)$, we deduce:

$$\frac{1}{2}\|\mathbf{SHu}^{k+1} - \mathbf{g}\|_2^2 + \frac{\beta_t^k}{2}\|\mathbf{D}_h \mathbf{u}^{k+1} - \mathbf{t}^{k+1} + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_2^2 \tag{21}$$
$$+ \frac{\beta_s^k}{2}\|\mathbf{D}_v \mathbf{u}^{k+1} - \mathbf{s}^{k+1} + \frac{\boldsymbol{\lambda}_s^k}{\beta_s^k}\|_2^2 \leq \frac{1}{2}\|\mathbf{SHu}^k - \mathbf{g}\|_2^2 + 2\mu N.$$

Since the all the terms on the LHS of (21) are nonnegative, the following inequality holds:

$$\frac{1}{2}\|\mathbf{SHu}^{k+1} - \mathbf{g}\|_2^2 \leq \frac{1}{2}\|\mathbf{SHu}^k - \mathbf{g}\|_2^2 + 2\mu N \leq \ldots \tag{22}$$
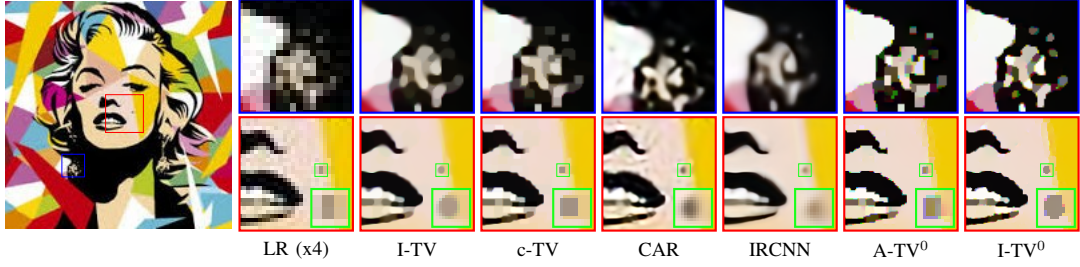$$\leq \frac{1}{2}\|\mathbf{SHu}^0 - \mathbf{g}\|_2^2 + 2\mu N k$$

Figure 8: Detail-preserving image cartoonisation by SR models. For I-TV, c-TV, A-TV^0 and I-TV^0 the regularisation parameters are chosen as μ: 0.08, 0.05, 0.02, 0.02, respectively.

From (21) and by the sub-additivity property of the square root we can also derive the following inequality:

$$\|\mathbf{D}_h\mathbf{u}^{k+1} - \mathbf{t}^{k+1} + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_2 \leq \sqrt{\frac{1}{\beta_t^k}}\|\mathbf{S}\mathbf{H}\mathbf{u}^0 - \mathbf{g}\|_2 + \sqrt{4\mu N\frac{k}{\beta_t^k}}$$

(23)

Step 3: We show that the sequences $\mathbf{D}_h\mathbf{u}^k$ and $\mathbf{D}_v\mathbf{u}^k$ are Cauchy sequences, hence they converge. We prove this for $\mathbf{D}_h\mathbf{u}^k$, the proof for $\mathbf{D}_v\mathbf{u}^k$ is identical.

$$\|\mathbf{D}_h\mathbf{u}^{k+1} - \mathbf{D}_h\mathbf{u}^k\|_2 \leq$$
$$\leq \|\mathbf{D}_h\mathbf{u}^{k+1} - \mathbf{t}^{k+1} + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_2 + \|\mathbf{D}_h\mathbf{u}^k - \mathbf{t}^{k+1} + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_2.$$

By assumption **A.1** applied on the RHS of (23) we deduce:

$$\|\mathbf{D}_h\mathbf{u}^{k+1} - \mathbf{t}^{k+1} + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}\|_2 \to 0, \tag{24}$$

which, combined with (20) and (23) entails that $\mathbf{D}_h\mathbf{u}^k$ is a Cauchy sequence. Hence it converges to a point $\mathbf{t}^*$. Similarly, $\mathbf{D}_v\mathbf{s}^k$ converges to a point $\mathbf{s}^*$.

Step 4: We prove now the convergence of the sequences $\mathbf{t}^k$ and $\mathbf{D}_h\mathbf{u}^k$. By writing (11) as:

$$\frac{\boldsymbol{\lambda}_t^{k+1}}{\beta_t^k} = \mathbf{D}_h\mathbf{u}^{k+1} - \mathbf{t}^{k+1} + \frac{\boldsymbol{\lambda}_t^k}{\beta_t^k}, \tag{25}$$

and from (24) we deduce that $\frac{\|\boldsymbol{\lambda}_t^{k+1}\|_2}{\sqrt{\beta_t^k}} \to 0$. By monotonicity of the (β_t^k) we then deduce that $\frac{\|\boldsymbol{\lambda}_t^k\|_2}{\sqrt{\beta_t^k}} \to 0$. Hence:

$$\|\mathbf{D}_h\mathbf{u}^{k+1} - \mathbf{t}^{k+1}\|_2 \leq \frac{\|\boldsymbol{\lambda}_t^{k+1}\|_2 + \|\boldsymbol{\lambda}_t^k\|_2}{\sqrt{\beta_t^k}},$$

where both quantities on the RHS tend to 0 as $k \to \infty$. Therefore, by the uniqueness of the limit, $\mathbf{t}^k \longrightarrow \mathbf{t}^*$ and $\mathbf{D}_h\mathbf{u}^k \longrightarrow \mathbf{t}^*$.

Step 5: We can now prove convergence of the sequence $(\mathbf{u}^k)$. For simplicity, let us define the quantities $\mathbf{A} := \mathbf{S}\mathbf{H}$ and $\mathbf{M}_k := \frac{1}{\beta_t^k}\mathbf{A}^T\mathbf{A} + \mathbf{D}_h^T\mathbf{D}_h + \frac{\beta_s^k}{\beta_t^k}\mathbf{D}_v^T\mathbf{D}_v$, for every k. By **A.2**, we observe that the matrix $\mathbf{M}_k$ is invertible for all k and that the optimality condition of (10) reads:

$$\mathbf{M}_k\mathbf{u}^k = \mathbf{D}_h^T(\mathbf{t}^{k+1} - \frac{\boldsymbol{\lambda}_t^k}{\beta^k}) + \frac{\beta_s^k}{\beta_t^k}\mathbf{D}_v^T(\mathbf{s}^{k+1} - \frac{\boldsymbol{\lambda}_s^k}{\beta^k}) + \frac{1}{\beta_t^k}\mathbf{A}^T\mathbf{g}.$$

Since $\mathbf{t}^{k+1} \to \mathbf{t}^*$, $\mathbf{s}^{k+1} \to \mathbf{s}^*$, $\frac{\boldsymbol{\lambda}_t^k}{\beta_t^k} \to 0$, $\frac{\boldsymbol{\lambda}_s^k}{\beta_s^k} \to 0$, and by Assumptions **A.1** and **A.2**, we have that $\frac{1}{\beta_t^k}\mathbf{A}^T\mathbf{g} \to \mathbf{0}$ so that the RHS converges pointwise to $\mathbf{z}^* = \mathbf{D}_h^T t^* + c\mathbf{D}_v^T s^*$. Additionally, the sequence $\mathbf{M}_k^{-1}$ converges pointwise to $\mathbf{M}^*$. We thus have that $\mathbf{u}^k = \mathbf{M}_k^{-1}\mathbf{M}_k\mathbf{u}^k \to \mathbf{M}^*\mathbf{z}^* := \mathbf{u}^*$.

We now want to show that $\mathbf{t}^* = \mathbf{D}_h\mathbf{u}^*$ and, similarly, that $\mathbf{s}^* = \mathbf{D}_v\mathbf{u}^*$. We show the details only for the former case. By the triangle inequality we get:

$$\|\mathbf{t}^* - \mathbf{D}_h\mathbf{u}^*\|_2 \leq \|\mathbf{t}^* - \mathbf{D}_h\mathbf{u}^k\|_2 + \|\mathbf{D}_h\mathbf{u}^k - \mathbf{D}_h\mathbf{u}^*\|_2$$
$$\leq \|\mathbf{t}^* - \mathbf{D}_h\mathbf{u}^k\|_2 + \|\mathbf{D}_h\|_2\|\mathbf{u}^k - \mathbf{u}^*\|_2,$$

where both terms tend to $\mathbf{0}$ since $\mathbf{D}_h\mathbf{u}^k \to \mathbf{t}^*$ and $\mathbf{u}^k \to \mathbf{u}^*$. ∎

B. Proof of Theorem (3)

Proof: The proof of Theorem (3) follows the same steps as the previous one. The only main difference in it is the definiton of $\mathbf{M}_k$, which reads in this case:

$$\mathbf{M}_k := \frac{1}{\beta^k}\mathbf{A}^T\mathbf{A} + \mathbf{D}^T\mathbf{D} = \frac{1}{\beta^k}\mathbf{A}^T\mathbf{A} + \mathbf{D}_h^T\mathbf{D}_h + \mathbf{D}_v^T\mathbf{D}_v.$$

By proceeding similarly as above the conclusion holds. ∎

REFERENCES

[1] M. Storath, A. Weinmann, and L. Demaret, "Jump-sparse and sparse recovery using Potts functionals," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3654–3666, 2014.

[2] C. G. Galbraith and J. A. Galbraith, "Super-resolution microscopy at a glance," *Journal of Cell Science*, vol. 124, no. 10, pp. 1607–1611, 2011.

[3] R. M. Willett, I. Jermyn, R. D. Nowak, and J. Zerubia, "Wavelet-Based Superresolution in Astronomy," in *ADASS XIII*, vol. 314, 2004, p. 107.

[4] Y. Kato, D. Deguchi, T. Takahashi, I. Ide, and H. Murase, "Low resolution QR-code recognition by applying super-resolution using the property of qr-codes," in *ICDAR 2011*, vol. IEEE ICDAR, pp. 992–996.

[5] F. Ling, Y. Zhang, G. M. Foody, X. Li, X. Zhang, S. Fang, W. Li, and Y. Du, "Learning-based superresolution land cover mapping," *IEEE Trans. Geosci. Remote Sens*, vol. 54, no. 7, pp. 3794–3810, 2016.

[6] J. Wickham, C. Homer, J. Vogelmann, A. McKerrow, R. Mueller, N. Herold, and J. Coulston, "The multi-resolution land characteristics (MRLC) consortium—20 years of development and integration of USA national land cover data," *Remote Sens.*, vol. 6, pp. 7424–7441, 2014.

[7] S. Chaudhuri, *Super-Resolution Imaging.* Springer US, 2001, vol. International Series in Engineering and Computer Science, no. 632.

[8] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Process.*, vol. 128, 2016.

[9] M. Storath, A. Weinmann, J. Frikel, and M. Unser, "Joint image reconstruction and segmentation using the Potts model," *Inverse Probl.*, vol. 31, no. 2, pp. 025 003, 29, 2015.

[10] M. Storath, D. Rickert, M. Unser, and A. Weinmann, "Fast segmentation from blurred data in 3D fluorescence microscopy," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4856–4870, 2017.

[11] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," *ACM Trans. Graph (SIGGRAPH Asia)*, 2011.

[12] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

[14] S. Gazagnes, E. Soubies, and L. Blanc-Féraud, "High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation," in *ISBI 2017*, 2017, pp. 28–31.

[15] T. F. Chan, M. K. Ng, A. C. Yau, and A. M. Yip, "Superresolution image reconstruction using fast inpainting algorithms," *Appl. Comput. Harmon. A.*, vol. 23, no. 1, pp. 3 – 24, 2007.

[16] S. J. O. A. Marquina, "Image super-resolution by TV-regularization and bregman iteration," *J. Sci. Comput.*, vol. 37, pp. 367–382, 2008.

[17] M. Tao, J. Yang, and B. He, "Alternating direction algorithms for total variation deconvolution in image reconstruction," *TR0918, Department of Mathematics, Nanjing University*, 2009.

[18] B. Gao, F. Sun, Y. Tong, and S. Xu, "Solving total-variation image super-resolution problems via proximal symmetric alternating direction methods," *J. Inequal. Appl.*, vol. 2016, no. 1, p. 197, 2016.

[19] W. Yao, J. Shen, Z. Guo, J. Sun, and B. Wu, "A total fractional-order variation model for image super-resolution and its SAV algorithm," *J. Sci. Comput.*, vol. 82, no. 3, pp. 1–18, 2020.

[20] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.

[21] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.

[22] W. Sun and Z. Chen, "Learned image downscaling for upscaling using content adaptive resampler," *IEEE Trans. Image Process.*, vol. 29, pp. 4027–4040, 2020.

[23] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proceedings of the IEEE CVPR*, 2017, pp. 3929–3938.

[24] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *IEEE GlobalSIP 2013*, pp. 945–948.

[25] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[26] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. A.*, vol. 27, no. 3, pp. 265–274, 2009.

[27] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for ℓ_2 - ℓ_2 problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, 2016.

[28] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, 2004.

[29] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, no. 8, pp. 1016–1022, 1979.

[30] T. Zhang, "Multi-stage convex relaxation for learning with sparse regularization," in *Adv. Neural. Inf. Process. Syst.*, 2009, pp. 1929–1936.

[31] F. Chen, L. Shen, and B. W. Suter, "Computing the proximity operator of the ℓ_p norm with $0 < p < 1$," *IET Signal Processing*, vol. 10, no. 5, pp. 557–565, 2016.

[32] D. Sage, T.-A. Pham, H. Babcock, T. Lukes, T. Pengo, J. Chao, R. Velmurugan, A. Herbert, A. Agrawal, S. Colabrese, A. Wheeler, A. Archetti, B. Rieger, R. Ober, G. M. Hagen, J.-B. Sibarita, J. Ries, R. Henriques, M. Unser, and S. Holden, "Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software," *Nature methods*, vol. 16, no. 5, p. 387—395, May 2019.

[33] M. Schwendy, R. E. Unger, and S. H. Parekh, "EVICAN–a balanced dataset for algorithm development in cell and nucleus segmentation," *Bioinformatics*, 2020.

[34] J. Cihlar, "Land cover mapping of large areas from satellites: status and research priorities," *Int. J. Remote Sens.*, vol. 21, no. 6-7, pp. 1093–1114, 2000.

[35] K. Malkin, C. Robinson, L. Hou, R. Soobitsky, J. Czawlytko, D. Samaras, J. Saltz, L. Joppa, and N. Jojic, "Label super-resolution networks," in *ICLR*, 2018.

[36] A. E. Maxwell, T. A. Warner, B. C. Vanderbilt, and C. A. Ramezan, "Land cover classification and feature extraction from national agriculture imagery program (NAIP) orthoimagery: A review," *ISPRS*, vol. 83, no. 11, pp. 737–747, 2017.

[37] R. Kemker, C. Salvaggio, and C. Kanan, "High-resolution multispectral dataset for semantic segmentation," *arXiv preprint arXiv:1703.01918*, 2017.

[38] P. Cascarano, F. Corsini, S. Gandolfi, E. L. Piccolomini, E. Mandanici, L. Tavasci, and F. Zama, "Super-resolution of thermal images using an automatic total variation based method," *Remote Sens.*, vol. 12, no. 10, p. 1642, 2020.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015.* Cham: Springer International Publishing, 2015, pp. 234–241.

Pasquale Cascarano received his master degree in Mathematics in 2018 at University of Bologna. He is currently a Ph.D. student in Applied Mathematics at the University of Bologna (IT). His research focuses on variational and deep learning methods for imaging inverse problems.

Luca Calatroni completed his Ph.D. in Applied Mathematics in 2015 as part of the Cambridge Image Analysis research group (UK). He was then a Marie Skłowdoska-Curie research fellow at the University of Genova (Italy) and *Lecteur Hadamard* FMJH fellow at the École Polytechnique (France). From October 2019, he is permanent CNRS researcher at the I3S laboratory in Sophia Antipolis (France) within the Morpheme research group. His research focuses on variational methods and non-smooth optimisation algorithms for imaging and vision.

Elena Loli Piccolomini is Professor of Numerical Analysis at the University of Bologna. Her research topics are regularisation methods for inverse problems in imaging, with particular focus on medical imaging and tomographic image reconstruction.