Nonconvex nonsmooth optimization via convex–nonconvex majorization–minimization

DOI: http://doi.org/10.1007/s00211-016-0842-x

Numerische Mathematik manuscript No.
(will be inserted by the editor)

Nonconvex Nonsmooth Optimization via Convex-Nonconvex Majorization-Minimization

A. Lanza · S. Morigi · I. Selesnik · F. Sgallari

the date of receipt and acceptance should be inserted later

Abstract The class of majorization-minimization algorithms is based on the principle of successively minimizing upper bounds of the objective function. Each upper bound, or surrogate function, is locally tight at the current estimate, and each minimization step decreases the value of the objective function. We present a majorization-minimization approach based on a novel convex-nonconvex upper bounding strategy for the solution of a certain class of nonconvex nonsmooth optimization problems. We propose an efficient algorithm for minimizing the (convex) surrogate function based on the alternating direction method of multipliers. A preliminary convergence analysis for the proposed approach is provided. Numerical experiments show the effectiveness of the proposed method for the solution of nonconvex nonsmooth minimization problems.

Keywords non-convex non-smooth optimization · majorization-minimization · regularization · alternating directions method of multipliers.

A. Lanza
Department of Mathematics, University of Bologna, Bologna, Italy.
E-mail: alessandro.lanza2@unibo.it

S. Morigi
Department of Mathematics, University of Bologna, Bologna, Italy.
E-mail: serena.morigi@unibo.it

I. Selesnik
Department of Electrical and Computer Engineering, New York University, Brooklyn, NY 11201 E-mail: selesi@nyu.edu.

F. Sgallari
Department of Mathematics, University of Bologna, Bologna, Italy.
E-mail: fiorella.sgallari@unibo.it

1 Introduction

This paper is concerned with the computation of approximate solutions $x^* \in \mathbb{R}^n$ of nonconvex nonsmooth minimization problems of the form

$$\arg\min_{x\in\mathbb{R}^n} \mathcal{J}(x), \qquad \mathcal{J}(x) := \frac{1}{2}\,\|Ax-b\|_2^2 \,+\, \sum_{i=1}^{s} \mu_i \phi\left((Lx)_i; a_i\right) \,, \qquad (1)$$

where $\|v\|_2$ and v_i denote the ℓ_2 norm and the i-th element of a vector v, respectively, $b \in \mathbb{R}^m$ is the vector of observed data, $\mu_i > 0$, $i = 1, \ldots, s$, represent the adaptive counterpart of the classical regularization parameter and control the trade-off between fidelity to the observations and regularity in the solutions x^* of (1), $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{s\times n}$ can be either the identity operator ($m = n$ and/or $s = n$) or generic linear operators. The so-called penalty or potential functions $\phi(\cdot; a_i) : \mathbb{R} \to \mathbb{R}$ are nonconvex nonsmooth functions parameterized by the scalar parameters $a_i > 0$ which control the degree of nonconvexity of $\phi(\cdot; a_i)$ and will be referred to as the *concavity parameters*. The aim/effect of using such nonconvex penalty functions is to promote sparsity of the vector Lx^* in the solutions of (1). The high flexibility of model (1) provided by the adaptive parameters μ_i, a_i can be reduced to a unique scalar value for μ and a if not strictly required.

The general minimization problem (1) encompasses a wide variety of problems that have been extensively studied in many different research areas, including numerical linear algebra [5, 50], image restoration [29, 46], pattern recognition [19, 33], and compressed sensing [8, 20]. Different choices of the linear operators A and L yield a variety of popular models that have been successfully used in many research and application fields. For instance, model (1) with $A \in \mathbb{R}^{m\times n}$, $m < n$, and L the identity operator can be used to compute sparse solutions of undetermined linear systems; when A is the identity operator and L is a linear operator representing discrete finite difference approximations of first- or second-order derivatives, model (1) can be used for denoising signals or images corrupted by additive Gaussian noise; when A is a sampling operator, model (1) can be applied for compressed sensing.

The functional $\mathcal{J}(x)$ in (1) is given by the sum of a convex smooth (quadratic) fidelity term and a nonconvex nonsmooth regularization term. Hence, $\mathcal{J}(x)$ is surely nonsmooth but can be convex or nonconvex depending on A, L, μ_i and a_i. In fact, in case that A has full column rank, i.e. A^TA is invertible, the quadratic fidelity term is strongly convex and its positive second-order derivatives holds the potential for compensating the negative second-order derivatives in the regularization term. The idea of constructing and then optimizing convex functionals containing nonconvex (sparsity-promoting) regularization terms, referred to as Convex-NonConvex (CNC) strategy, was first introduced by Blake and Zisserman [6] in the context of Graduated Non Convexity (GNC) and by Nikolova [37] for the denoising of binary images. The CNC approach has very recently been explored by Selesnick and others for different purposes, see [12, 18, 30, 32, 41, 46, 47] for more details. The attractiveness of such CNC

approach resides in its ability to promote sparsity more strongly than using convex regularization while at the same time maintaining the convexity of the optimization problem, so that well-known reliable convex minimization approaches can be used to compute the (unique) solution.

A first contribution of this paper is the derivation of conditions that ensure the functional $\mathcal{J}$ in problem (1) is convex – despite the regularization term being nonconvex. However, these conditions can be quite restrictive. Hence, in order to strongly promote sparsity, we are still interested in solving problem (1) when $\mathcal{J}$ is not convex. As a second contribution we thus allow $\mathcal{J}$ to be nonconvex, so that the regularizer can better approximate the ℓ_0 "norm", which is known to strongly promote sparsity. In this case, the CNC approach is not applied directly as in [12,18,30,32,41,46,47]. In fact, we present an iterative approach to solve problem (1) when the functional $\mathcal{J}$ is not convex, that leverages the CNC strategy at each iteration. In particular, to obtain an approximate solution of problem (1) we use the Majorization-Minimization (MM) procedure, which consists in replacing the original problem (1) by a sequence of simpler problems. Specifically, the k-th iteration of the standard MM approach applied to the solution of (1) consists of two main computational steps: a majorization step which generates a so-called *surrogate* function majorizing (i.e., upper bounding) the objective function in (1), and a minimization step which computes the minimizer of this surrogate function:

S1) Generate a surrogate function $\mathcal{S}(x, x^{(k)}) : \mathbb{R}^n \to \mathbb{R}$, majorizing $\mathcal{J}(x)$ at $x^{(k)}$

S2) Compute the next iterate by solving

$$x^{(k+1)} = \arg\min_{x \in \mathbb{R}^n} S(x, x^{(k)}) \tag{2}$$

We propose a non-standard class of surrogate functions with the aim that they more accurately approximate the objective function $\mathcal{J}$ as compared to the standard approaches. Consequently, we expect to obtain faster convergence and more robust convergence to the global minimizer. Commonly used majorizers proposed in literature include quadratic [17, 42] and piecewise linear [2, 9] functions, which we will refer to as Q-MM and L-MM, respectively. In contrast to these majorizers (which are convex), in this paper we propose to majorize the regularization term by a nonconvex function, with the constraint that the total surrogate function $S(x, x^{(k)})$ in (2) be strictly convex in x. We will refer to this method as Convex-NonConvex Majorization-Minimization (CNC-MM).

Summarizing, the key contributions of this paper are as follows:

a) Derivation of sufficient conditions for strict convexity of the cost functional $\mathcal{J}(x)$ in (1) which generalize results presented in [47], [46], [41];
b) Proposal of a novel CNC-MM strategy applied to the solution of nonconvex nonsmooth optimization problems of the form (1);
c) Analysis of convergence of the proposed CNC-MM approach applied to (1);
d) Proposal of an efficient algorithm for minimizing the (convex) surrogate function based on the Alternating Direction Method of Multipliers (ADMM).

In particular, we address the case where A in problem (1) has full column rank, i.e., A^TA is invertible. The more general case (which includes compressed sensing, ill-conditioned deconvolution, super-resolution, and tomography) will be addressed in future work and will build upon the results herein for the full column rank case. The case considered in this paper has applications in denoising using total variation [30,46], wavelets [18], and the short-time Fourier transform [12].

This paper is organized as follows. In Section 2 we formally define the class of nonconvex nonsmooth sparsity-promoting penalty functions used in the considered model (1) and then give some popular examples. In Section 3 we derive sufficient conditions for strict convexity of model (1). In Section 4 and Section 5 we illustrate in detail how the majorization and minimization steps are carried out in our CNC-MM proposal, respectively. More precisely, in Section 4 we outline the proposed strategy for generating CNC surrogate functions majorizing the nonconvex nonsmooth original functional, whereas in Section 5 we present an efficient algorithm for minimizing the CNC surrogate functions based on the ADMM. Convergence analysis is carried out in Section 6. Numerical examples evaluating the performance of the proposed approach are presented in Section 7 and, finally, conclusions are drawn in Section 8.

1.1 Related work

The proposed approach builds upon and complements other work to solve nonconvex problems of the form (1), in particular majorization-minimization (MM) [21, 28] and graduated nonconvexity (GNC) [6, 38, 40].

Other types of algorithms aiming to solve nonsmooth linear inverse problems of this form include iteratively reweighted least squares (IRLS) [17], iteratively reweighted ℓ_1 (IRL1) [2, 9], FOCUSS-type algorithms [35, 42]. Algorithms developed for the nonconvex case include [13,48], and [10,11], and new proximal algorithms [14, 43].

Since the proposed approach calls for the solution of a sequence of convex problems, some convex optimizations algorithm can be utilized. A few such algorithms are the iterative shrinkage/thresholding algorithm (ISTA/FISTA) [5, 22], proximal methods [4, 15, 16], and the ADMM [1, 7, 25].

2 Nonconvex penalty functions

We denote the sets of non-negative and positive real numbers as $\mathbb{R}_+ := \{\, t \in \mathbb{R} : t \geqslant 0\}$ and $\mathbb{R}_+^* := \{\, t \in \mathbb{R} : t > 0\}$, respectively. We consider parameterized penalty functions $\phi(\cdot\,; a) : \mathbb{R} \to \mathbb{R}$ such that for any value of the parameter $a \in \mathbb{R}_+^*$ the following assumptions are satisfied:

A1) $\phi(\cdot\,; a) \in \mathcal{C}^0(\mathbb{R})$ (ϕ continuous in t on $\mathbb{R}$)

A2) $\phi(-t; a) = \phi(t; a) \;\; \forall\, t \in \mathbb{R}_+^*$ (ϕ even in t)

A3) $\phi(\cdot;a) \in \mathcal{C}^2(\mathbb{R}_+^*)$ (ϕ twice continuously differentiable in t on $\mathbb{R}_+^*$)
A4) $\phi'(t;a) > 0 \;\; \forall\, t \in \mathbb{R}_+^*$ (ϕ strictly increasing in t on $\mathbb{R}_+^*$)
A5) $\phi''(t;a) \leqslant 0 \;\; \forall\, t \in \mathbb{R}_+^*$ (ϕ concave in t on $\mathbb{R}_+^*$)
*A6) $\sup_{t\in\mathbb{R}_+^*} \phi'(t;a) < +\infty, \; \inf_{t\in\mathbb{R}_+^*} \phi''(t;a) > -\infty$ (ϕ'/ϕ'' bounded from above/below)

A6) $\phi(0;a) = 0, \; \sup_{t\in\mathbb{R}_+^*} \phi'(t;a) = 1, \; \inf_{t\in\mathbb{R}_+^*} \phi''(t;a) = -a$ (ϕ, ϕ', ϕ'' normalization)

We denoted by $\phi'(t;a)$ and $\phi''(t;a)$ the first-order and second-order derivatives of ϕ with respect to the variable t, respectively, while in the following $\phi'_a(t;a)$ indicates the mixed second-order derivative with respect to t and the parameter a. Assumptions A1)–A5) are quite standard and encompass a wide class of nonsmooth nonconvex sparsity-promoting penalty functions [23, 24]. Assumption *A6), which corresponds to imposing boundedness of first-order and second-order derivatives of the penalty function $\phi(\cdot;a)$, is mandatory when constructing CNC functionals (see, e.g., [12,30,41,46,47]). In fact, if the second-order derivative of the penalty function goes to $-\infty$ at any point in the domain, there is no possibility to compensate it by the positive but bounded second-order derivatives of the convex quadratic fidelity term. In particular, we notice that the popular ℓ_p quasi-norm with $0 < p < 1$, namely $\phi(t;p) = |t|^p$, even if it has been successfully applied, e.g., in image restoration [31], satisfies assumptions A1)–A5) but not *A6), since $\phi'(0^+;p) = +\infty, \; \phi''(0^+;p) = -\infty$. As a consequence, such a penalty function does not allow for applying the CNC strategy. Efficient strategies for the solution of (1) where the ℓ_p quasi-norm is used can be found in [29, 44].

We notice that, without loss of generality, assumption *A6) can be replaced with A6) which can be easily obtained by scaling and normalization. Assumption A6) represents a useful normalization of the penalty functions which has been used, e.g., in [12, 18]. In particular, from the assumption on $\phi''(t;a)$ it follows that the parameter a represents a scalar indicator of the "degree of nonconvexity" of the penalty function ϕ, thus justifying the name concavity parameter.

The following assumptions A7)–A9) are proposed in this paper to allow for constructing CNC majorizing functions, as it will be illustrated in detail in Section 4.1:

A7) $\phi'(t;\cdot), \, \phi''(t;\cdot) \in \mathcal{C}^1(\mathbb{R}_+^*) \;\; \forall\, t \in \mathbb{R}_+^*$ (ϕ', ϕ'' differentiable in a on $\mathbb{R}_+^*$)
A8) $\phi'_a(t;a) < 0 \;\; \forall\, t \in \mathbb{R}_+^*$ (ϕ' non-increasing in a on $\mathbb{R}_+^*$)
A9) $\dfrac{\phi''(t;a_1)}{\phi'(t;a_1)} \geqslant \dfrac{\phi''(t;a_2)}{\phi'(t;a_2)} \;\; \forall\, t \in \mathbb{R}_+^*, \quad \forall\, a_1 < a_2$

We notice that A9) represents a monotonic relationship between the parameter a and the ratio between the first-order ϕ' and second-order ϕ'' derivatives.

Finally, we notice that any penalty function satisfying assumptions A1)–A9) tends to the absolute value function $f(t) = |t|$ when a approaches zero. Hence, we set:

$$\phi(t;0) := |t| \, . \tag{3}$$

	$\phi_{\log}$	ϕ_{rat}	ϕ_{atan}
ϕ	$\frac{\log(1+a\lvert t\rvert)}{a}$	$\frac{\lvert t\rvert}{1+a\lvert t\rvert/2}$	$\frac{\mathrm{atan}\left(\frac{1+2a\lvert t\rvert}{\sqrt{3}}\right)-\frac{\pi}{6}}{a\sqrt{3}/2}$
ϕ'	$\frac{\mathrm{sign}(t)}{1+a\lvert t\rvert}$	$\frac{\mathrm{sign}(t)}{(1+a\lvert t\rvert/2)^2}$	$\frac{\mathrm{sign}(t)}{1+a\lvert t\rvert+a^2t^2}$
ϕ''	$-\frac{a}{(1+a\lvert t\rvert)^2}$	$-\frac{a}{(1+a\lvert t\rvert/2)^3}$	$-\frac{a\,(1+2a\lvert t\rvert)}{\left(1+a\lvert t\rvert+a^2t^2\right)^2}$
ϕ'''	$\frac{2a^2\mathrm{sign}(t)}{(1+a\lvert t\rvert)^3}$	$\frac{3a^2\mathrm{sign}(t)}{2(1+a\lvert t\rvert/2)^4}$	$\frac{6a^3t(1+a\lvert t\rvert)}{\left(1+a\lvert t\rvert+a^2t^2\right)^3}$
$\lim_{t\to\pm\infty}\phi$	$+\infty$	$\frac{2}{a}$	$\frac{2}{a}\ \frac{\pi\sqrt{3}}{9}$

Table 1 Examples of penalty functions satisfying assumptions A1)–A9).

According to (3), the ℓ_1-norm penalty is recovered as a special case of the parameterized penalty function $\phi(\cdot;a)$ when $a=0$.

In Table 1 we report three examples of (sparsity-promoting) penalty functions, referred to as $\phi_{\log}$, ϕ_{rat}, ϕ_{atan}, which satisfy all the above assumptions A1)–A9) and will be used in the paper. These functions have been considered variously in e.g. [9, 24, 39]. In Table 1 we also report the associated first, second and third-order derivatives. In Figure 1 we show the plots of the penalty functions $\phi_{\log}$, ϕ_{rat}, ϕ_{atan} with a varying value of the concavity parameter ($a=2,3,4$).

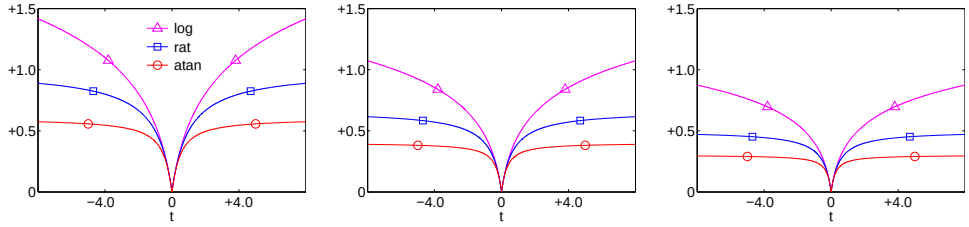

Fig. 1 Plots of the penalty functions $\phi_{\log}(t;a)$, $\phi_{\mathrm{rat}}(t;a)$, $\phi_{\mathrm{atan}}(t;a)$ defined in Table 1, for different values of the concavity parameter a: $a=2$ (left), $a=3$ (center), $a=4$ (right).

3 Convexity conditions for $\mathcal{J}(x)$

In this section, we investigate convexity of functional $\mathcal{J}(x)$ in (1). More precisely, we seek to find sufficient conditions on the linear operators $A\in\mathbb{R}^{m\times n}$ and $L\in\mathbb{R}^{s\times n}$ and on the parameters $\mu_i\in\mathbb{R}_+^*$ and $a_i\in\mathbb{R}_+^*$, $i=1,\dots,s$, to ensure that $\mathcal{J}(x)$ is strictly convex in its entire domain $x\in\mathbb{R}^n$. We notice that, in practice, the linear operators A and L are generally prescribed by the specific application considered for model (1), so that the derived convexity conditions can be regarded as constraints on the free parameters μ_i and a_i.

The aim of the investigation is twofold. First, we could impose constraints on the parameters μ_i and a_i such that $\mathcal{J}(x)$ is strictly convex. This would allow

us to directly apply the CNC approach to model (1). This will be considered in the experimental section, but is not the main contribution of this paper. A more significant benefit is obtained by pushing the functional $\mathcal{J}(x)$ beyond the derived convexity limits, and, consequently, by applying a CNC-MM optimization upper bounding it by a sequence of surrogate functions which instead satisfy constraints on the penalty parameters and are thus strictly convex. This represents the main goal of this paper.

Let us first introduce two useful lemmas: the first one follows easily and does not need a demonstration, whereas the proof of the latter can be found in [41].

Lemma 1 *Let $Z \in \mathbb{R}^{d_1 \times d_1}$, $Y \in \mathbb{R}^{d_2 \times d_2}$, $X \in \mathbb{R}^{d_2 \times d_1}$, $d_1 \leqslant d_2$, be matrices such that $Z = X^T Y X$. Then, a necessary and sufficient condition for Z being symmetric positive definite is that X has full column rank and Y is symmetric positive definite.*

Lemma 2 *Let $\phi(\,\cdot\,;a) : \mathbb{R} \to \mathbb{R}$ be a function satisfying assumptions A1)–A6) in Section 2 with $a \in \mathbb{R}^*_+$ and let $u(\,\cdot\,;a) : \mathbb{R} \to \mathbb{R}$ be the function defined by*

$$u(t;a) \;:=\; \phi(t;a) - |t| \;. \tag{4}$$

Then, the following statements hold:

1) $u(\cdot;a) \in \mathcal{C}^2(\mathbb{R}) \qquad \forall\, a \in \mathbb{R}^*_+ \,,$

2) $u''(t;a) \in [\,-a\,,\,0\,] \quad \forall\, a \in \mathbb{R}^*_+, \;\; \forall t \in \mathbb{R} \,.$

Based on the above lemmas, in the proposition below we give sufficient conditions for convexity of $\mathcal{J}(x)$.

Proposition 1 *Let $\mathcal{J} : \mathbb{R}^n \to \mathbb{R}$ be the functional defined in (1), with $\phi : \mathbb{R} \to \mathbb{R}$ a penalty function satisfying assumptions A1)–A6) in Section 2, $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{s \times n}$ and $\mu_i \in \mathbb{R}^*_+$, $a_i \in \mathbb{R}^*_+$ $\forall\, i \in \{1, \ldots, s\}$. Then, $\mathcal{J}$ is strictly convex if the following two conditions both hold:*

$$1) \;\; \ker\left\{A^T A\right\} = \{0\} \,, \tag{5}$$

$$2) \;\; \mu_i a_i < \rho(A,L) \quad \forall\, i \in \{1, \ldots, s\} \,, \tag{6}$$

where the positive scalar ρ, depending on matrices A and L, is defined as:

$$\rho(A,L) \;:=\; \frac{\sigma^2_{A,\min}}{\sigma^2_{L,\max}} \,, \tag{7}$$

with $\sigma_{A,\min}$ and $\sigma_{L,\max}$ denoting the minimum singular value of matrix A and the maximum singular value of matrix L, respectively.

The proof is provided in the Appendix.

Our proposal requires the knowledge of the values of $\sigma_{A,\min}$ and $\sigma_{L,\max}$ for computing the scalar value ρ in (7), which, in many popular applications, can be derived by explicit formulas. This is the case for a few interesting models of the form (1) that make use of the following well-know matrices A and L:

- $A = I_n$, $L = D_1, D_2$, where D_1, D_2 represent the dicretization of the first- or second- order derivatives, for example signal/image denoising,
- $A = I_n$, $L := W$ where W represents an orthogonal basis or an over-complete dictionary, which satisfies the tight frame condition, i.e., $L^T L = rI_n$, $r > 0$, for signal/image reconstruction (analysis approach),
- $A := W^{-1}$, $L = I_n$ for the sparse reconstruction (synthesis approach).

In a general case where no explicit expressions for $\sigma_{A,\min}$ and $\sigma_{L,\max}$ are available, efficient numerical procedures can be used for their accurate estimation [3].

In case that the functional $\mathcal{J}(x)$ in the original optimization problem (1) satisfies both conditions 1) and 2) in Proposition 1, i.e. $\mathcal{J}(x)$ is strictly convex, then there is no reason to apply the proposed CNC-MM approach. In fact, in this convex case at the first iteration the CNC-MM algorithm would generate a surrogate function which is $\mathcal{J}(x)$ itself, thus reducing the CNC-MM procedure to a single iteration yielding the global minimizer; this will be clarified later in Section 4.2.

Usefulness of the proposed CNC-MM strategy is for solving optimization problems of the form (1) when the functional $\mathcal{J}(x)$ satisfies (5) but violates the convexity limits in (6). Hence, in the rest of the paper we will assume that functional $\mathcal{J}(x)$ in the original problem (1) satisfies the following two hypothesis:

$$\text{H1)} \quad \ker\left\{A^T A\right\} = \{0\}\,, \tag{8}$$

$$\text{H2)} \quad \mu_i a_i = \bar{\tau}\,\rho(A, L) \quad \forall\, i \in \{1, \ldots, s\}\,, \quad \bar{\tau} \in \left]\,1\,, +\infty\,\right[\,. \tag{9}$$

Hypotheses H1) is necessary for allowing the proposed CNC-MM approach to generate CNC (strictly convex) surrogate functions according to conditions 1) and 2) in Proposition 1. Hypotheses H2) formalizes the fact that $\mathcal{J}(x)$ must be beyond the convexity limits, such that the CNC-MM approach can be usefully applied. In particular, since there is no a priori reason to impose different degrees of nonconvexity for the s penalty terms in the regularizer of $\mathcal{J}(x)$, a unique scalar coefficient $\bar{\tau} > 1$ is used which simultaneously sets the degree of concavity of all the regularization terms.

4 The CNC-MM algorithm: Majorization by CNC surrogates

In this section, we illustrate in detail the majorization step of the proposed CNC-MM approach. More precisely, we outline the strategy used at any iteration k of the CNC-MM algorithm for generating a CNC (strictly convex)

surrogate function $S(x, x^{(k)})$ majorizing the nonconvex nonsmooth functional $\mathcal{J}(x)$ in (1) at the current iterate $x^{(k)}$. Provided that it is not necessary to majorize the quadratic strictly convex (A has full column rank) fidelity term of $\mathcal{J}(x)$, and that the nonconvex regularization term consists of the sum of s occurrences of the same nonconvex ϕ function, majorization of $\mathcal{J}(x)$ reduces to majorization of the penalty function ϕ. Hence, in subsection 4.1 we illustrate how to majorize a nonconvex nonsmooth penalty function ϕ satisfying the assumptions outlined in Section 2 by another nonconvex nonsmooth (surrogate) function having a smaller degree of concavity. Then, in Subsection 4.2 we show how to use these results, together with the convexity conditions derived in Section 3, in order to construct a CNC majorizer of the entire functional $\mathcal{J}(x)$.

4.1 Majorization of nonconvex penalty functions

In the following Definition 1 we formally introduce the concept of tangent majorant function [27], [26]. Then, in Proposition 2 we recall some known results about quadratic and piecewise linear majorization, respectively. In Proposition 3 we present a novel nonconvex majorization procedure for nonconvex penalty functions satisfying assumptions A1)–A9) in Section 2.

Definition 1 Let $\mathcal{F}(x) : \mathbb{R}^n \to \mathbb{R}$ be a continuous not necessarily smooth function. Then the function $\mathcal{M}(x, v) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is said to be a tangent majorant for $\mathcal{F}(x)$ at any $v \in \mathbb{R}^n$ if and only if all the following conditions hold:

$$\begin{aligned}
&1. \quad \mathcal{M}(v,v) = \mathcal{F}(v) \qquad \forall\, v \in \mathbb{R}^n\,,\\
&2. \quad \nabla_x \mathcal{M}(v,v) = \nabla_x \mathcal{F}(v) \quad \forall\, v \in \mathbb{R}^n,\\
&3. \quad \mathcal{M}(x,v) \geqslant \mathcal{F}(x) \qquad \forall\, v \in \mathbb{R}^n\,, \quad \forall\, x \in \mathbb{R}^n\,,
\end{aligned}$$

where ∇_x denotes the ordinary gradient operator for majorization points v where the functions are smooth, and the set of all possible directional derivatives for singular points.

Proposition 2 *Let $\phi\,(\,\cdot\,;a)\!:\!\mathbb{R}\!\to\!\mathbb{R}$ be a function satisfying assumptions A1)–A6) in Section 2 with given $a \in \mathbb{R}_+^*$. Then, the function $q(\,\cdot\,,\,\cdot\,;a)\!:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ defined by*

$$q(t,v;a) := w_q\,\frac{t^2}{2} + c_q, \quad \text{with} \quad w_q := \frac{1}{v}\,\phi'(v;a),\ c_q := \phi(v;a) - \frac{v}{2}\,\phi'(v;a), \tag{10}$$

is a (quadratic, strictly convex) tangent majorant for $\phi\,(\,\cdot\,;a)$ at any $v \in \mathbb{R}\setminus\{0\}$.

The function $l(\,\cdot\,,\,\cdot\,;a)\!:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ defined by

$$l(t,v;a) := w_l\,|t| + c_l, \quad \text{with} \quad w_l := \phi'(|v|;a),\ c_l := \phi(v;a) - v\,\phi'(v;a), \tag{11}$$

	$\phi_{\log}$	$\phi_{\rm rat}$	$\phi_{\rm atan}$
w_q	$\dfrac{1}{\lvert v\rvert(1+a\lvert v\rvert)}$	$\dfrac{1}{\lvert v\rvert(1+a\lvert v\rvert/2)^2}$	$\dfrac{1}{\lvert v\rvert(1+a\lvert v\rvert+a^2v^2)}$
w_l	$\dfrac{1}{1+a\lvert v\rvert}$	$\dfrac{1}{(1+a\lvert v\rvert/2)^2}$	$\dfrac{1}{1+a\lvert v\rvert+a^2v^2}$
w_m	$\dfrac{1+a_m\lvert v\rvert}{1+a\lvert v\rvert}$	$\left(\dfrac{2+a_m\lvert v\rvert}{2+a\lvert v\rvert}\right)^2$	$\dfrac{1+a_m\lvert v\rvert+a_m^2v^2}{1+a\lvert v\rvert+a^2v^2}$

Table 2 Explicit expressions for the weights w_q, w_l, w_m associated with quadratic, piecewise linear and nonconvex majorization, respectively, of penalty functions $\phi_{\log}$, $\phi_{\rm rat}$, $\phi_{\rm atan}$ defined in Table 1.

is a (piecewise linear, convex, nonsmooth) tangent majorant for $\phi(\,\cdot\,;a)$ *at any* $v\in\mathbb{R}$.

Proposition 3 *Let* $\phi(\,\cdot\,;a):\mathbb{R}\to\mathbb{R}$ *be a function satisfying assumptions A1)–A9) in Section 2 with given* $a\in\mathbb{R}_+^*$. *Then, any function* $m(\,\cdot\,,\,\cdot\,;a_m):\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ *belonging to the* a_m*-parameterized family defined by*

$$m(t,v;a_m) \;:=\; w_m\,\phi(t;a_m)+c_m\,,\quad a_m\in\,]\,0,a\,]\,, \tag{12}$$

$$\text{with}\quad w_m:=\frac{\phi'(v;a)}{\phi'(v;a_m)}\,,\qquad c_m:=\phi(v;a)-w_m\,\phi(v;a_m)\,, \tag{13}$$

is a (nonconvex nonsmooth) tangent majorant for $\phi(\,\cdot\,;a)$ *at any* $v\in\mathbb{R}$, *that is:*

$$m(v,v;a_m) \;=\; \phi(v;a) \qquad \forall\, a_m\in\,]0,a],\,\forall\, v\,\in\mathbb{R} \tag{14}$$

$$m_t(v,v;a_m) \;=\; \phi'(v;a) \qquad \forall\, a_m\in\,]0,a],\,\forall\, v\,\in\mathbb{R}\setminus\{0\} \tag{15}$$

$$m_t(0^{\pm},0;a_m) \;=\; \phi'(0^{\pm};a) \quad \forall\, a_m\in\,]0,a], \tag{16}$$

$$m(t,v;a_m) \;\geqslant\; \phi(t;a) \qquad \forall\, a_m\in\,]0,a],\,\forall\, v\,\in\mathbb{R},\,\forall\, t\in\mathbb{R} \tag{17}$$

The proof is provided in the Appendix.

In Table 2 we report explicit formulas for the computation of weights w_q, w_l, w_m associated with quadratic, piecewise linear and nonconvex majorization, respectively, of the penalty functions $\phi_{\log}$, $\phi_{\rm rat}$, $\phi_{\rm atan}$ defined in Table 1.

In Figure 2 we depict the majorants at a given abscissa $v=0.5$ of the nonconvex function $\phi_{\log}(t;a)$ with parameter value $a=2.5$. In particular, in Figure 2(left) we show the unique convex quadratic majorant $q(t,v)$ calculated as in (10), the unique convex piecewise linear majorant $l(t,v)$ defined in (11), and one among the infinitely many nonconvex majorants $m(t,v;a_m)$ given in (12), namely the one obtained by using $a_m=1.0<a$. It can be noticed how the nonconvex majorant approximates the function $\phi_{\log}$ better than both the linear and the quadratic majorants. In Figure 2(right) we report three different nonconvex majorants obtained by using $a_m=0.5,1.0,1.5$, respectively.

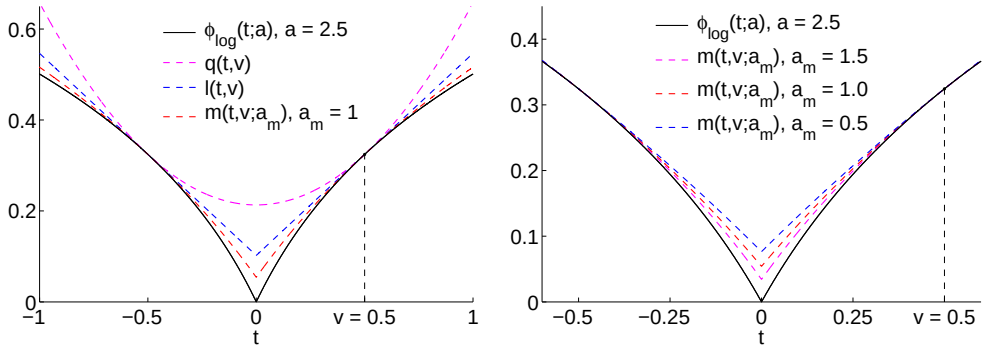

Fig. 2 Majorants at $v = 0.5$ of the nonconvex function $\phi_{\log}(t; a)$. Left: quadratic, piecewise linear and nonconvex majorants. Right: three nonconvex majorants with different value of the parameter $a_m < a$.

We end this subsection by reporting a corollary of Proposition 3 that will be useful when formalizing CNC majorization of the entire functional $\mathcal{J}(x)$.

Corollary 1 *Let $\phi(\,\cdot\,; a) : \mathbb{R} \to \mathbb{R}$ be a function satisfying assumptions A1)–A9) in Section 2. Then, for any $a \in \mathbb{R}_+^*$ and $v \in \mathbb{R}$, to be considered constant parameters, the majorization weight function $w_m(a_m; a, v) : \mathbb{R}_+^* \to \mathbb{R}$, considered as function of the independent variable a_m, and defined in (13) has the following properties:*

$$\begin{aligned} w_m(a; a, v) = 1\,, \;\; w_m(a_m; a, v) > 0 \quad \forall\, a_m \in \mathbb{R}_+^*\,, \\ w_m(\cdot; a, v) \in \mathcal{C}^1(\mathbb{R}_+^*)\,, \;\; w'_m(a_m; a, v) > 0 \quad \forall\, a_m \in \mathbb{R}_+^*\,. \end{aligned} \tag{18}$$

Proof By substituting a for a_m in (13), it is clear that $w_m(a; a, v) = 1$. The second property in (18), namely $w_m(a_m; a, v) > 0$, follows directly from assumption A4) in Section 2. Then, we notice that the majorization weight defined in (13), when regarded as a function $w_m(a_m; a, v)$ of the parameter a_m, is differentiable in a_m due to assumption A7) and its first-order derivative $w'_m(a_m; a, v)$ is as follows:

$$w'_m(a_m; a, v) \;=\; -\,\phi'(v; a)\; \frac{\phi'_{a_m}(v; a_m)}{\big(\phi'(v; a_m)\big)^2}\,. \tag{19}$$

Due to assumptions A4) and A8), the first-order derivative in (19) is positive for any $a_m \in \mathbb{R}_+^*$, thus concluding the proof.

4.2 Majorization of $\boldsymbol{\mathcal{J}(x)}$

At each iteration k of the MM approach formalized in (2) and applied to the solution of the considered problem (1), the surrogate function $S(x, x^{(k)})$ to be minimized for computing the next iterate $x^{(k+1)}$ is generated by independently

majorizing all the occurrences of the penalty functions $\phi(\cdot; a_i)$, $i = 1, \ldots, s$, in the regularization term of the original functional $\mathcal{J}(x)$ in (1). More precisely, based on the results presented in the previous section, the surrogate functions $S_q(x, x^{(k)})$, $S_l(x, x^{(k)})$, $S_m(x, x^{(k)})$ obtained by using quadratic, piecewise linear and the proposed nonconvex majorization of $\phi(\cdot; a_i)$ according to (10), (11), (12), respectively, take the forms:

$$S_q(x, x^{(k)}) = \frac{1}{2}\|Ax - b\|_2^2 + \sum_{i=1}^{s}\left[\mu_i\, w_{q,i}^{(k)}\, \frac{(Lx)_i^2}{2}\right] + C_q^{(k)}\,, \tag{20}$$

$$S_l(x, x^{(k)}) = \frac{1}{2}\|Ax - b\|_2^2 + \sum_{i=1}^{s}\left[\mu_i\, w_{l,i}^{(k)}\, |(Lx)_i|\right] + C_l^{(k)}\,, \tag{21}$$

$$S_m(x, x^{(k)}) = \frac{1}{2}\|Ax - b\|_2^2 + \sum_{i=1}^{s}\left[\mu_i\, w_{m,i}^{(k)}\, \phi\big((Lx)_i; a_{m,i}^{(k)}\big)\right] + C_m^{(k)}\,, \tag{22}$$

where $C_q^{(k)}$, $C_l^{(k)}$, $C_m^{(k)}$ are constants not depending on x and the majorization weights, following (10), (11), (13), are as follows:

$$w_{q,i}^{(k)} = \frac{\phi'(v_i^{(k)}; a_i)}{v_i^{(k)}}\,, \qquad w_{l,i}^{(k)} = \phi'(|v_i^{(k)}|; a_i)\,, \qquad w_{m,i}^{(k)} = \frac{\phi'(v_i^{(k)}; a_i)}{\phi'(v_i^{(k)}; a_{m,i}^{(k)})}\,, \tag{23}$$

$$\text{where} \quad v_i^{(k)} := (Lx^{(k)})_i\,. \tag{24}$$

Results in Proposition 2 guarantee that the surrogate functions $S_q(x, x^{(k)})$ and $S_l(x, x^{(k)})$ in (20)–(21) are tangent majorants at $x^{(k)}$ of the original functional $\mathcal{J}(x)$ in (1). By introducing the three $s \times s$ diagonal matrices M and $W_q^{(k)}, W_l^{(k)}$ having on the main diagonal the regularization parameters and the quadratic and piecewise linear majorization weights, that is

$$M := \mathrm{diag}(\mu_1, \ldots, \mu_s)\,, \tag{25}$$

$$W_q^{(k)} := \mathrm{diag}(w_{q,1}^{(k)}, \ldots, w_{q,s}^{(k)})\,, \qquad W_l^{(k)} := \mathrm{diag}(w_{l,1}^{(k)}, \ldots, w_{l,s}^{(k)})\,, \tag{26}$$

the surrogate functions $S_q(x, x^{(k)})$ and $S_l(x, x^{(k)})$ in (20)–(21) can be rewritten in the following more compact and more popular forms:

$$S_q(x, x^{(k)}) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\left\| M^{1/2} {W_q^{(k)}}^{1/2} Lx \right\|_2^2 + C_q^{(k)}\,, \tag{27}$$

$$S_l(x, x^{(k)}) = \frac{1}{2}\|Ax - b\|_2^2 + \left\| MW_l^{(k)} Lx \right\|_1 + C_l^{(k)}\,. \tag{28}$$

The surrogate function forms in (27)–(28) motivate the names *iterated reweighted least squares* (or IRL2) and *iterated reweighted least absolute deviations* (or IRL1), alternatively used for the MM approach with quadratic and piecewise linear majorization, respectively. Since the regularization parameters μ_i are positive by assumption and the majorization weights in (23) are positive by construction, the diagonal matrices $M, W_q^{(k)}, W_l^{(k)}$ in (25)–(26) are positive

definite. Therefore, under the (mild) assumption that $\ker\{A^TA\} \cap \ker\{L^TL\} = \{0\}$, the surrogate functions $S_q(x, x^{(k)})$ and $S_l(x, x^{(k)})$ in (27)–(28) are strictly convex and can be reliably minimized at any iteration of the MM approach.

For what concerns the proposed nonconvex majorization strategy, Proposition 3 guarantees that $S_m(x, x^{(k)})$ in (22) is a tangent majorant at $x^{(k)}$ of the original functional $\mathcal{J}(x)$ in (1) provided that the penalty function ϕ satisfies assumptions A1)–A9) outlined in Section 2 and that the concavity parameters $a_{m,i}^{(k)}$ in (22) are chosen so as to satisfy the following constraints:

$$a_{m,i}^{(k)} \in \,]\,0, a_i\,] \qquad \forall\, i \,\in \{1, \ldots, s\}\,, \tag{29}$$

where a_i are the predefined concavity parameters in the original cost functional $\mathcal{J}(x)$ in (1). However, satisfying conditions (29) does not guarantee that the surrogate function $S_m(x, x^{(k)})$ is convex. In fact, $S_m(x, x^{(k)})$ in (22) has the same form of functional $\mathcal{J}(x)$ in (1), for which convexity conditions are derived in Section 3. In particular, based on Proposition 1, the surrogate function $S_m(x, x^{(k)})$ in (22) is strictly convex if the following two conditions are both satisfied:

$$\begin{cases} \ker\{A^TA\} = \{0\}\,, \\ \mu_i\, w_{m,i}^{(k)}\, a_{m,i}^{(k)} < \rho(A, L) \quad \forall\, i \,\in \{1, \ldots, s\}\,, \end{cases} \tag{30}$$

where the positive scalar $\rho(A, L)$ is defined in (7), Proposition 1. The first condition in (30) is satisfied since it coincides with the first hypotheses H1) in (8) on the original functional $\mathcal{J}(x)$. The second condition in (30) is more complex and will be dealt with in the rest of this section.

From the second hypotheses H2) on the original functional $\mathcal{J}(x)$ formalized in (9), we can derive the following expressions for the regularization parameters μ_i:

$$\mu_i \;=\; \frac{\overline{\tau}\,\rho(A, L)}{a_i}\,, \quad i \in \{1, \ldots, s\}\,. \tag{31}$$

Substituting (31) in the second condition of (30), we obtain that the surrogate function in (22) is strictly convex if:

$$a_{m,i}^{(k)}\; w_{m,i}^{(k)}\big(a_{m,i}^{(k)}; a_i, v_i^{(k)}\big) \;<\; \gamma_i := \frac{a_i}{\overline{\tau}} \quad \forall\, i \,\in \{1, \ldots, s\}\,. \tag{32}$$

where, for clarity, we made explicit all the dependencies of the majorization weights $w_{m,i}^{(k)}$ according to their definition in (23)–(24). We remark that the right-hand sides γ_i of inequalities (32) are positive scalar constants, while the left-hand sides are nonlinear functions of the concavity parameters $a_{m,i}^{(k)}$, which represent the free-parameters we aim to select for constructing a strictly convex surrogate function, i.e. the unknowns of convexity conditions (32).

In order to solve inequalities in (32), we introduce the following lemma, where for simplicity of notations we drop the subscripts i and the superscripts (k).

Lemma 3 *Let $a \in \mathbb{R}^*_+$, $v \in \mathbb{R}$ be given constants, let $\phi(\,\cdot\,;a)\colon \mathbb{R} \to \mathbb{R}$ be a function satisfying assumptions A1)–A9) in Section 2 and let $\varrho(\cdot;a,v) : \mathbb{R}^*_+ \to \mathbb{R}$ be the function of the independent variable a_m defined as:*

$$\varrho(a_m;a,v) \,:=\, a_m\, w_m(a_m;a,v)\,, \qquad w_m(a_m;a,v) := \frac{\phi'(v;a)}{\phi'(v;a_m)}\,, \qquad a_m > \,0\,. \tag{33}$$

Then, for any $a \in \mathbb{R}^*_+$, $v \in \mathbb{R}$ and any $\gamma \in \mathbb{R}^*_+$, the constrained (nonlinear) inequality in the unknown a_m defined by:

$$\varrho(a_m;a,v) \,<\, \gamma\,, \qquad a_m > \,0\,, \tag{34}$$

admits the solution interval

$$a_m \in \,]\,0\,,\, \varrho^{-1}(\gamma;a,v)\,[\,, \tag{35}$$

where $\varrho^{-1}(\cdot;a,v)\colon \mathbb{R}^*_+ \to \mathbb{R}^*_+$, representing the inverse of function $\varrho(\cdot;a,v)$ in (33), associates any $\gamma \in \mathbb{R}^*_+$ with the (unique positive) solution of the following nonlinear equation in the unknown a_m:

$$a_m\, \phi'(v;a) - \gamma\, \phi'(v;a_m) \,=\, 0\,, \qquad a_m > \,0\,. \tag{36}$$

Moreover, for any $a \in \mathbb{R}^*_+$, $v \in \mathbb{R}$ the function $\varrho^{-1}(\cdot;a,v)$ has the following property:

$$\gamma \in \,]\,0\,,\, a\,[\;\;\Longrightarrow\;\; \varrho^{-1}(\gamma;a,v) \in \,]\,0\,,\, a\,[\,. \tag{37}$$

Proof It follows from Corollary 1 that for any $a \in \mathbb{R}^*_+$, $v \in \mathbb{R}$, the function $\varrho(\cdot;a,v)$ of the independent variable a_m defined in (33) has the following properties:

$$\begin{gathered} \varrho(0^+;a,v) = 0\,,\;\; \varrho(a;a,v) = a\,, \\ \varrho(\cdot;a,v) \in \mathcal{C}^1(\mathbb{R}^*_+)\,,\;\; \varrho'(a_m;a,v) > 0 \quad \forall\, a_m \in \mathbb{R}^*_+\,. \end{gathered} \tag{38}$$

In fact, the first three among the properties in (38) follow directly from definition (33) and from the first three properties of the majorization weight function $w_m(\cdot;a,v)$ outlined in (18), Corollary 1. Moreover, the first-order derivative (with respect to a_m) of $\varrho(\cdot;a,v)$ in (33) is $\varrho'(a_m;a,v) = w_m(a_m;a,v) + a_m w'_m(a_m;a,v)$. Since $w_m, w'_m > 0$ for any $a_m \in \mathbb{R}^*_+$ by (18), the last property in (38) follows. According to (38), the function $\varrho(\cdot;a,v)$ of the independent variable a_m in (33) starts from 0 and is continuously differentiable and monotonically increasing on its entire domain $a_m \in \mathbb{R}^*_+$. Hence, $\varrho(\cdot;a,v)$ is bijective, admits inverse function $\varrho^{-1}(\cdot;a,v) : \mathbb{R}^*_+ \to \mathbb{R}^*_+$ and for any $\gamma \in \mathbb{R}^*_+$ the solutions of inequality (34) are given by (35).

We can finally summarize in the following result the conditions for S_m to be a strictly convex majorant.

Proposition 4 *Let $\mathcal{J}(x) : \mathbb{R}^n \to \mathbb{R}$ be the functional defined in* (1)*, with $\phi : \mathbb{R} \to \mathbb{R}$ a penalty function satisfying assumptions A1)–A9) in Section 2. Then, for any $x^{(k)} \in \mathbb{R}^n$, the surrogate function $S_m(x, x^{(k)})$ defined in* (22) *is a tangent majorant of $\mathcal{J}(x)$ at $x^{(k)}$ if the following conditions hold:*

$$a_{m,i}^{(k)} \in \,]\,0\,,\,a_i\,] \qquad \forall\, i \in \{1,\dots,s\}\,. \tag{39}$$

Moreover, the surrogate function $S_m(x, x^{(k)})$ in (22) *is a strictly convex tangent majorant of $\mathcal{J}(x)$ at $x^{(k)}$ if the following more stringent conditions hold:*

$$a_{m,i}^{(k)} \in \,]\,0\,,\,\varrho^{-1}(\gamma_i; a_i, v_i^{(k)})\,[\qquad \forall\, i \in \{1,\dots,s\}\,, \tag{40}$$

where $\varrho^{-1}(\gamma_i; a_i, v_i^{(k)})$ are given in Proposition (5), the (constant) scalars $\gamma_i \in \mathbb{R}_+^*$ and $v_i^{(k)} \in \mathbb{R}$ are defined in (32) and (24), respectively.

Proof Conditions (39) for the surrogate function S_m to be a majorant are obtained by Proposition 4.4. as derived in (29). By applying Lemma 3 to inequalities (32), we easily derive the conditions (40) for S_m to be strictly convex, for any $a_i \in \mathbb{R}_+^*$, $v_i^{(k)} \in \mathbb{R}$, $\gamma_i \in \mathbb{R}_+^*$. Moreover, since $\bar{\tau} > 1$ by hypotheses H2) in (9) on $\mathcal{J}(x)$, it follows from the definition of γ_i in (32) that $\gamma_i \in \,]\,0\,,\,a_i\,[\ \forall\, i \in \{1,\dots,s\}$. Hence, according to statement (37) in Lemma 3, we have that:

$$\varrho^{-1}(\gamma_i; a_i, v_i^{(k)}) \in \,]\,0\,,\,a_i\,[\quad \forall\, i \in \{1,\dots,s\}\,. \tag{41}$$

It follows from (41) that the intervals defined in (40) are nonempty proper subsets of the corresponding intervals in (29), that is the strict convexity conditions are more stringent than the majorization conditions.

We can thus conclude that the surrogate function $S_m(x, x^{(k)})$ in (22) is a strictly convex tangent majorant at $x^{(k)}$ of the original functional $\mathcal{J}(x)$ in (1) if the concavity parameters $a_{m,i}^{(k)}$ are all chosen so as to satisfy conditions (40). In practice, in order to construct the (strictly convex majorant) surrogate $S_m(x, x^{(k)})$, we need to select a specific value $\hat{a}_{m,i}^{(k)}$ inside the intervals (40) for each concavity parameter $a_{m,i}^{(k)}$. Analogously to what we did in hypothesis H2) in (9), since there is no a priori reason for imposing different degrees of convexity for the s terms in the regularizer of the surrogate function $S_m(x, x^{(k)})$ in (22), we introduce a unique scalar coefficient $\underline{\tau} \in]0, 1[$ aimed to simultaneously set the degree of convexity of all the regularization terms in $S_m(x, x^{(k)})$. In particular, the specific values are selected as follows:

$$\hat{a}_{m,i}^{(k)} = \varrho^{-1}(\underline{\tau}\gamma_i; a_i, v_i^{(k)})\,, \quad i \in \{1,\dots,s\}\,. \tag{42}$$

Proposition 5 *Let $\phi_{\log}$, ϕ_{rat}, ϕ_{atan} be the penalty functions defined in Table 1. Then, for any $a \in \mathbb{R}_+^*$, $v \in \mathbb{R}$, the function $\varrho^{-1}(\gamma; a, v)$ in (35) admits the following explicit expression:*

$$\varrho^{-1}(\gamma; a, v) = \begin{cases} \gamma & \text{if } v = 0 \\ \varrho_{\log}^{-1}(\gamma; a, v), \varrho_{\mathrm{rat}}^{-1}(\gamma; a, v), \varrho_{\mathrm{atan}}^{-1}(\gamma; a, v) & \text{if } v \neq 0\,,\ \phi = \phi_{\log}, \phi_{\mathrm{rat}}, \phi_{\mathrm{atan}} \end{cases}, \tag{43}$$

where:

$$\varrho_{\log}^{-1} = \frac{1}{|v|}\left(\sqrt{\frac{1}{4} + \gamma|v|(1 + a|v|)} - \frac{1}{2}\right) \tag{44}$$

$$\varrho_{\mathrm{rat}}^{-1} = \frac{1}{3|v|}\left(E + \frac{4}{E} - 4\right) \quad \text{with} \quad \begin{cases} E = \sqrt[3]{G + \sqrt{G^2 - 64}} \\ G = 8 + \dfrac{27}{2}\,\gamma|v|\,(2 + a|v|)^2 \end{cases} \tag{45}$$

$$\varrho_{\mathrm{atan}}^{-1} = \frac{1}{3|v|}\left(E - \frac{2}{E} - 1\right) \quad \text{with} \quad \begin{cases} E = \sqrt[3]{G + \sqrt{G^2 + 8}} \\ G = \dfrac{7}{2} + \dfrac{27}{2}\,\gamma|v|(1 + a|v| + a^2|v|^2) \end{cases} \tag{46}$$

Proof First, we notice that in case that $v = 0$ the equation in (36) leads to the unique solution $a_m = \varrho^{-1}(\gamma; a, v) = \gamma$. This follows from the fact that $w_m(a_m, a, 0) = 1$ for the penalty functions considered. Replacing into equation (36) the expression of $\phi'_{\log}$, ϕ'_{rat}, ϕ'_{atan} given in the second row of Table 1, we obtain, respectively:

$$\begin{aligned} \phi_{\log}: \quad & a_m^2|v| + a_m - \gamma\,(1 + a|v|) && = 0 \\ \phi_{\mathrm{rat}}: \quad & a_m^3 v^2 + 4\,a_m^2|v| + 4\,a_m - \gamma\,(2 + a|v|)^2 && = 0 \quad , a_m > 0\,. \\ \phi_{\mathrm{atan}}: \quad & a_m^3 v^2 + a_m^2|v| + a_m - \gamma\,(1 + a|v| + a^2 v^2) && = 0 \end{aligned}$$

By applying closed-form formulas for the zeros of quadratic and cubic equations, respectively, explicit expressions for a_m in (44), (45), (46) are obtained.

Finally, in case the surrogate function $S_m(x, x^{(k)})$ is a strictly convex tangent majorant of $\mathcal{J}(x)$ at $x^{(k)}$, then its expression in (22), by using (31), (32), and (42), can be rewritten as follows:

$$S_m(x, x^{(k)}) = \frac{1}{2}\,\|Ax - b\|_2^2 + \underline{\tau}\rho(A, L) \sum_{i=1}^{s} \frac{1}{\hat{a}_{m,i}^{(k)}}\,\phi\big((Lx)_i; \hat{a}_{m,i}^{(k)}\big) + C_m^{(k)}\,. \tag{47}$$

5 The CNC-MM algorithm: Minimization by ADMM

In this section, we illustrate the ADMM-based [7] iterative algorithm used at each iteration k of the proposed CNC-MM approach to compute the new iterate $x^{(k+1)}$ by minimizing the nonsmooth strictly convex surrogate function $S_m(x, x^{(k)})$ defined in (47). More precisely, we aim at solving the following unconstrained minimization problem:

$$y^* \;=\; \arg\min_{y\in\mathbb{R}^n} \left\{ \frac{1}{2\,\underline{\tau}\rho}\,\|Ay-b\|_2^2 \;+\; \sum_{i=1}^{s} \frac{1}{a_i}\;\phi\big((Ly)_i; a_i\big) \right\}, \tag{48}$$

where in (48) the constant $C_m^{(k)}$ in (47) has been omitted, the objective function has been divided by the positive constant $\underline{\tau}\rho$ and, to simplify notations, the superscripts (k) have been drop and the optimization variable y is used in place of x.

In order to apply ADMM for the solution of (48), we first resort to the variable splitting technique [1] and introduce the auxiliary variable $d \in \mathbb{R}^s$, such that problem (48) is reformulated into the following linearly constrained equivalent form:

$$\{\, y^*, d^* \} \;=\; \arg\min_{y\in\mathbb{R}^n} \left\{ \frac{1}{2\,\underline{\tau}\rho}\,\|Ay-b\|_2^2 \;+\; \sum_{i=1}^{s} \frac{1}{a_i}\;\phi(d_i; a_i) \right\} \quad \text{s.t.} \quad d = Ly\,. \tag{49}$$

The auxiliary variable d is aimed to transfer the terms $(Ly)_i$ in (48) out of the nonconvex nonsmooth penalty functions $\phi(\cdot; a_i)$. To solve (49), we define the augmented Lagrangian functional

$$\mathcal{L}(y, d; \lambda) = \frac{1}{2\,\underline{\tau}\rho}\,\|Ay-b\|_2^2 + \sum_{i=1}^{s} \frac{1}{a_i}\;\phi(d_i; a_i) - \langle\, \lambda\,, d - Ly \,\rangle + \frac{\beta}{2}\;\|\, d - Ly \,\|_2^2\,, \tag{50}$$

where $\beta > 0$ is a scalar penalty parameter and $\lambda \in \mathbb{R}^s$ is the vector of Lagrange multipliers associated with the system of linear constraints $d = Ly$ in (49). Solving (49) is thus equivalent to seeking for the solutions of the following saddle-point problem:

$$\begin{aligned} \text{Find} \quad & (y^*, d^*; \lambda^*) \;\in\; \mathbb{R}^n \times\; \mathbb{R}^s \times\; \mathbb{R}^s \\ \text{s.t.} \quad & \mathcal{L}\,(y^*, d^*; \lambda) \;\leqslant\; \mathcal{L}\,(y^*, d^*; \lambda^*) \;\leqslant\; \mathcal{L}\,(y, d; \lambda^*) \\ & \forall\,(y, d; \lambda) \;\in\; \mathbb{R}^n \times\; \mathbb{R}^s \times\; \mathbb{R}^s\,. \end{aligned} \tag{51}$$

Given the previously computed (or initialized for $j = 0$) vectors $y^{(j)}$ and $\lambda^{(j)}$, the j-th iteration of the ADMM iterative scheme [7] applied to the solution of (48) or, equivalently, to the saddle-point problem (50)–(51), reads as:

$$d^{(j+1)} \;\leftarrow\; \arg\min_{d\in R^s}\; \mathcal{L}(y^{(j)}, d; \lambda^{(j)}) \tag{52}$$

$$y^{(j+1)} \;\leftarrow\; \arg\min_{y\in R^n}\; \mathcal{L}(y, d^{(j+1)}; \lambda^{(j)}) \tag{53}$$

$$\lambda^{(j+1)} \;\leftarrow\; \lambda^{(j)} \,-\, \beta\,\big(\, d^{(j+1)} - Ly^{(j+1)}\,\big). \tag{54}$$

In the following we show in detail how to solve the two minimization subproblems (52) and (53) for the primal variables d and y, respectively, then we present the overall CNC-MM algorithm.

Solving the subproblem for *d*. Given $y^{(j)}$ and $\lambda^{(j)}$, the minimization subproblem for the variable d in (52) can be rewritten as follows:

$$\begin{aligned} d^{(j+1)} &\leftarrow \arg\min_{d\in R^s} \left\{ \sum_{i=1}^{s} \left[\frac{1}{a_i}\,\phi(d_i;a_i) \right] - \left\langle \lambda^{(j)}, d - Ly^{(j)} \right\rangle + \frac{\beta}{2} \left\| d - Ly^{(j)} \right\|_2^2 \right\} \\ &\leftarrow \arg\min_{d\in R^s} \sum_{i=1}^{s} \left[\frac{1}{a_i}\,\phi(d_i;a_i) + \frac{\beta}{2} \left(d_i - z_i^{(j)} \right)^2 \right] , \end{aligned} \tag{55}$$

where in (55) the minimized functional is written in component-wise form and $z^{(j)} \in \mathbb{R}^s$ denotes the constant (with respect to the variable d) vector defined as

$$z^{(j)} := Ly^{(j)} + \frac{1}{\beta}\,\lambda^{(j)} . \tag{56}$$

The minimization in (55) is equivalent to the following s independent scalar problems:

$$d_i^{(j+1)} \leftarrow \arg\min_{d_i\in\mathbb{R}} \left\{ \frac{1}{2} \left(d_i - z_i^{(j)} \right)^2 + \frac{1}{a_i\beta}\,\phi(d_i;a_i) \right\}, \quad i = 1,\ldots,s . \tag{57}$$

Recalling that the cost function minimized in (48) is strictly convex by construction, we would like the ADMM subproblems in (57) also to be strictly convex. To this purpose, in the first part of Proposition 6 below we give useful convexity conditions.

In particular, we notice that the cost functions in problems (57) have the same form as the function f defined in (58), where the constants $z_i^{(j)}$, $1/a_i\beta$, a_i and the optimization variables d_i in (57) correspond to b, μ, a and t in (58), respectively.

Proposition 6 *Let $\phi(\,\cdot\,;a):\mathbb{R}\to\mathbb{R}$ be a function satisfying assumptions A1)–A6) in Section 2 and $a, \mu \in \mathbb{R}_+^*$, $b \in \mathbb{R}$ be given generic constants. Then, the function $f : \mathbb{R} \to \mathbb{R}$ defined as*

$$f(t) := \frac{1}{2}\,(t-b)^2 + \mu\,\phi(t;a) \tag{58}$$

is strictly convex in the variable t if and only if the following condition holds:

$$\mu\, a < 1 . \tag{59}$$

In case that (59) holds, the proximity map $\operatorname{prox}_\phi^\mu : \mathbb{R} \to \mathbb{R}$ of function ϕ defined as:

$$\operatorname{prox}_\phi^\mu(b) := \arg\min_{t\in\mathbb{R}} f(t) , \quad b \in \mathbb{R} , \tag{60}$$

is given by

$$\text{prox}_{\phi}^{\mu}(b) \;=\; \begin{cases} 0 & \text{if } \; |b| \leqslant \mu \\ \text{sign}(b)\, t^* & \text{if } \; |b| > \mu \;, \end{cases} \tag{61}$$

where t^* is the unique solution of the following constrained nonlinear equation:

$$t + \mu\, \phi'(t;a) - |b| = 0\,, \qquad 0 < t < |b|\;. \tag{62}$$

The proof of convexity condition (59) can be found in [12], whereas the proximity operator can be found in [47].

According to (59), problems (57) are strictly convex if and only if the following conditions hold:

$$\frac{1}{a_i \beta}\, a_i \;<\; 1 \quad \forall\, i \in \{1, \ldots, s\} \quad \Longleftrightarrow \quad \beta \;>\; 1\,. \tag{63}$$

It follows from (63) that all the s problems in (57) are strictly convex if and only if the ADMM scalar penalty parameter β is greater than one. We remark that the value of the parameter β only affects the speed of convergence of the ADMM iterative scheme toward the global minimizer y^* of (48).

In case that (63) is satisfied, the unique solutions of the strictly convex problems in (57) can be computed based on the proximity maps introduced in the second part of Proposition 6. In Proposition 7 below, we report some useful results which allow for the efficient solution of (62) under suitable assumptions on the penalty function ϕ. In particular, according to the result in (66), in case that $\phi \in \mathcal{C}^3(\mathbb{R}_+^*)$, $\phi'''(t;a) \geqslant 0\; \forall\, t \in \mathbb{R}_+^*$, the solution t_i^* of (62) can be obtained as the limit point of a (quadratically convergent) Newton-Raphson iteration.

We notice that the previously introduced penalty functions $\phi_{\log}$, ϕ_{rat} and ϕ_{atan} all satisfy the above condition on the third-order derivative, as can be seen in the fourth row of Table 1, hence the Newton-Raphson procedure can be applied. However, for these penalty functions the solutions of problems (57) can be computed even more efficiently by using the closed-form formulas reported in the second part of Proposition 7. For instance, in case that ϕ is the $\phi_{\log}$ penalty function (which will be considered in the experimental section), the solutions t_i^* of the nonlinear equations in (62) can be determined by means of the following closed-form formula

$$t_i^* \;=\; \frac{1}{a_i} \left(E_i - \frac{1}{2} + \sqrt{\left(E_i + \frac{1}{2} \right)^2 - \frac{1}{\beta}} \right) \quad \text{with} \quad E_i \;=\; \frac{a_i}{2}\, |z_i^{(j)}|\;. \tag{64}$$

Proposition 7 *Let $\phi\,(\,\cdot\,;a) : \mathbb{R} \to \mathbb{R}$ be a function satisfying assumptions A1)–A6) in Section 2, let $a, \mu \in \mathbb{R}_+^*$, $b \in \mathbb{R}$ be given constants satisfying:*

$$\mu\, a < 1\,, \qquad \mu < |b|\,, \tag{65}$$

and let t^* denote the unique solution of the nonlinear equation in (62). Then:

- *In case that* $\phi \in \mathcal{C}^3(\mathbb{R}^*_+)$, $\phi'''(t;a) \geqslant 0 \ \forall\, t \in \mathbb{R}^*_+$, t^* *can be obtained as the limit point of the following (quadratically convergent) Newton-Raphson iterative scheme:*

$$t^{(0)} = |b|\,, \qquad t^{(k+1)} = \frac{t^{(k)}\phi''\big(t^{(k)};a\big) - \phi'\big(t^{(k)};a\big) + |b|/\mu}{\phi''\big(t^{(k)};a\big) + 1/\mu}\,. \tag{66}$$

- *In case that* ϕ *is one among the penalty functions* $\phi_{\log}$, $\phi_{\rm rat}$, $\phi_{\rm atan}$ *defined in Table 1,* t^* *can be obtained as the unique solution of the following nonlinear polynomial equations:*

$$\begin{aligned}
\phi_{\log}: &\qquad\qquad a\,t^2 + (1 - a\,|b|)\,t + \mu - |b| = 0 \\
\phi_{\rm rat}: &\ \frac{a^2}{4}\,t^3 + a\,(1 - \frac{a}{4}|b|)\,t^2 + (1 - a\,|b|)\,t + \mu - |b| = 0 \\
\phi_{\rm atan}: &\ a^2 t^3 + a\,(1 - a|b|)\,t^2 + (1 - a\,|b|)\,t + \mu - |b| = 0
\end{aligned} \tag{67}$$

under the constraint $0 < t < |b|$.

Proof Let us define the function $g : \mathbb{R}^*_+ \to \mathbb{R}$ as

$$g(t) := t + \mu\,\phi'(t;a) - |b|\,, \quad t > 0\,, \tag{68}$$

such that (62) can be rewritten as follows: $g(t) = 0,\ 0 < t < |b|$. It is easy to demonstrate that the function g in (68) has the following properties:

$$\begin{gathered}
g \in \mathcal{C}^2(\mathbb{R}^*_+), \qquad g(0^+) = \mu - |b| < 0, \qquad g(|b|) = \mu\,\phi'(|b|;a) > 0, \\
g'(t) = 1 + \mu\,\phi''(t;a) > 0 \ \ \forall\, t > 0, \qquad g''(t) = \mu\,\phi'''(t;a) \geqslant 0 \ \ \forall\, t > 0\,.
\end{gathered}$$

Hence, since $g(t)$ is monotonically increasing and convex the Newton-Raphson iterative scheme applied to the solution of $g(t) = 0$, with initial guess $t^{(0)} = |b|$ is guaranteed to converge with quadratic rate toward the unique root t^* of $g(t)$ in the interval $]\,0, |b|\,[$. The scheme reads as follows:

$$t^{(k+1)} = t^{(k)} - \frac{g\big(t^{(k)}\big)}{g'\big(t^{(k)}\big)} = t^{(k)} - \frac{t^{(k)} + \mu\,\phi'\big(t^{(k)};a\big) - |b|}{1 + \mu\,\phi''\big(t^{(k)};a\big)}, \tag{69}$$

thus proving (66).

To derive (67), it is sufficient to substitute in the nonlinear equation (62) the expressions of $\phi'_{\log}$, $\phi'_{\rm rat}$, $\phi'_{\rm atan}$ given in Table 1, thus obtaining, respectively:

$$\begin{aligned}
\phi_{\log}: &\quad t - |b| + \frac{\mu}{1 + a\,t} = 0 \\
\phi_{\rm rat}: &\quad t - |b| + \frac{\mu}{1 + a\,t + a^2 t^2/4} = 0 \quad , \ 0 < t < |b|\,. \\
\phi_{\rm atan}: &\quad t - |b| + \frac{\mu}{1 + a\,t + a^2 t^2} = 0
\end{aligned}$$

We notice that the denominators of the above fractional terms are positive numbers. Hence, simple algebraic manipulations yield (67).

Solving the subproblem for y. Given $\lambda^{(j)}$ and $d^{(j+1)}$, and recalling the definition of the augmented Lagrangian functional in (50), the minimization subproblem for y in (53) can be rewritten as follows:

$$y^{(j+1)} \leftarrow \arg\min_{y \in R^n} \left\{ \frac{1}{2\,\underline{\tau}\rho} \|Ay - b\|_2^2 + \left\langle \lambda^{(j)}, Ly \right\rangle + \frac{\beta}{2} \left\| d^{(j+1)} - Ly \right\|_2^2 \right\}, \tag{70}$$

where constant terms have been omitted. We notice that (70) is a quadratic minimization problem whose first-order optimality conditions lead to:

$$\left(\frac{1}{\underline{\tau}\rho\,\beta} A^T A + L^T L \right) y \;=\; \frac{1}{\underline{\tau}\rho\,\beta} A^T b + L^T \left(d^{(j+1)} - \frac{1}{\beta} \lambda^{(j)} \right) . \tag{71}$$

The $n \times n$ linear system in (71) is solvable if the coefficient matrix has full rank, that is if the following condition holds:

$$\ker\{A^T A\} \,\cap\, \ker\{L^T L\} = \{0\} . \tag{72}$$

Since we are assuming that $\ker\{A^T A\} = \{0\}$ – see hypothesis H1) in (8) – it follows that (72) holds for any matrix L. Hence, the quadratic cost functional in (70) is strictly convex and its global minimizer $y^{(j+1)}$ can be obtained by computing the unique solution of (71).

Provided that the penalty parameter $\beta > 0$ is kept fixed during iterations, the coefficient matrix in (71) is symmetric positive definite and does not change with iterations (neither the inner ADMM iterations, nor the outer MM iterations). Hence, the linear system in (71) can be solved quite efficiently by the iterative (preconditioned) Conjugate Gradient method or by computing the Cholesky factorization once for all, depending on the size of the problem.

Moreover, in case the matrices A and L have a particular structured form, the linear system can be solved even more efficiently. For example, when $A^T A$ and $L^T L$ are circulant matrices, the coefficient matrix in (71) can be diagonalized by the discrete Fourier transform (FFT implementation).

The CNC-MM algorithm. To summarize previous results, in Algorithm 1 we report the main computational steps of the overall proposed CNC-MM iterative approach.

To compare the performance of our proposal with the popular Q-MM and L-MM alternative approaches based on the construction of the quadratic and piecewise linear surrogate functions $S_q(x, x^{(k)})$ and $S_l(x, x^{(k)})$ defined in (27)–(28), respectively, we report here some information about their minimization.

For what concerns the minimization of $S_q(x, x^{(k)})$ in (27), it reduces to the solution of the following linear system of normal equations:

$$\left(A^T A + L^T {W_q^{(k)}}^{1/2} M {W_q^{(k)}}^{1/2} L \right) x \;=\; A^T b . \tag{73}$$

Analogously to the linear system in (71), (73) is solvable thanks to hypotheses H1) in (8) and the coefficient matrix in (73) is symmetric positive definite due to diagonal matrices M and $W_q^{(k)}$ being positive definite. However, unlike

Algorithm 1 CNC-MM approach applied to the solution of problem (1)

inputs: $\phi : \mathbb{R} \to \mathbb{R}$ satisfying A1)–A9) in Section 2

$A \in \mathbb{R}^{m\times n}$ satisfying H1) in (8) and $L \in \mathbb{R}^{s\times n} \Rightarrow \rho(A, L) > 0$ by (7)

$\mu_i, a_i > 0$ satisfying H2) in (9) with $\bar{\tau} > 1$

$b \in \mathbb{R}^m$ (observed data), $x^{(0)} \in \mathbb{R}^n$ (initial guess)

parameters: $0 < \underline{\tau} < 1$ (majorization step, suggested value $\underline{\tau} = 0.99$)

$\beta > 1$ (minimization step, suggested value $\beta = 10$)

1: **for** $k = 0, 1, 2, \ldots$ until convergence **do**

- **Majorization step:** generate the surrogate $\mathcal{S}_m(x, x^{(k)})$ defined in (47):

2: $v^{(k)} \leftarrow Lx^{(k)}$

3: $\hat{a}_{m,i}^{(k)} \leftarrow \varrho^{-1}\big(\underline{\tau}\, a_i / \bar{\tau}; a_i, v_i^{(k)}\big)$, $i = 1, \ldots, s$, with ϱ^{-1} defined in (43)–(46)

- **Minimization step:** compute $x^{(k+1)} = \arg\min_{x\in\mathbb{R}^n} \mathcal{S}_m(x, x^{(k)})$ by ADMM:

4: $y^{(0)} \leftarrow x^{(k)}, \quad \lambda^{(0)} \leftarrow \lambda^{(j+1)}$ if $k > 0$ (0 if $k = 0$)

5: **for** $j = 0, 1, 2, \ldots$ until convergence **do**

6: given $y^{(j)}, \lambda^{(j)}$, compute $d^{(j+1)}$ by (56), (61), (60)

7: given $d^{(j+1)}, \lambda^{(j)}$, compute $y^{(j+1)}$ by solving (71)

8: given $y^{(j+1)}, d^{(j+1)}, \lambda^{(j)}$, compute $\lambda^{(j+1)}$ by (54)

9: **end for** j

10: $x^{(k+1)} \leftarrow y^{(j+1)}$

11: **end for** k

output: $x^{(k+1)} \in \mathbb{R}^n$ (approximate local/global minimizer of $\mathcal{J}(x)$)

(71), the coefficient matrix in (73) varies during the MM iterations due to the majorization weights matrix $W_q^{(k)}$ and is not diagonalizable by fast transforms such as Fourier, sine and cosine transforms. Hence, (73) is typically solved by the iterative CG method.

As far as the piecewise linear surrogate $S_l(x, x^{(k)})$ in (28) is concerned, its minimization can be easily carried out by suitably adapting the ADMM procedure illustrated above for the minimization of $S_m(x, x^{(k)})$. In particular, the minimization of $S_l(x, x^{(k)})$ can be rewritten in a form similar to (48) where the function ϕ is replaced by the absolute value function. The only significant difference in the ADMM scheme is in the subproblem for the variable d, which becomes inherently convex and allows a closed form solution [49] based on a well known soft thresholding operator.

We finally remark that the suggested values of the input parameters $\underline{\tau}$ and β are derived from numerical experiments.

6 Convergence Analysis

In this section we analyze convergence of the proposed CNC-MM approach, whose main computational steps are given in Algorithm 1, when applied to the solution of nonsmooth nonconvex optimization problems of the form (1) under hypothesis H1)–H2) in (8)–(9).

First, we characterize the objective function $\mathcal{J}(x)$ in (1) and the surrogate function $\mathcal{S}_m(x, x^{(k)})$ defined in (47).

Definition 2 A convex (not necessarily differentiable) function $f(x)$ is said to be δ-strongly convex if and only if there exists a constant $\delta > 0$, called the modulus of strong convexity of $f(x)$, such that the function $f(x) - \frac{\delta}{2}\|x\|_2^2$ is convex.

Proposition 8 *Under hypothesis H1)–H2) in (8)–(9), the objective function $\mathcal{J}(x)$ in (1) and the surrogate function $\mathcal{S}_m(x, x^{(k)})$ in (47) are proper, continuous (hence, lower semi-continuous), bounded from below and coercive functions. Moreover, $\mathcal{S}_m(x, x^{(k)})$ is δ-strongly convex with modulus of strong convexity*

$$\delta \;=\; \sigma_{A,min}^2\,(1 - \underline{\tau}) \;. \tag{74}$$

Proof Both $\mathcal{J}$ and $\mathcal{S}_m$ are clearly proper functions. Moreover, since the penalty function ϕ is continuous and bounded from below by zero and the quadratic fidelity term is coercive, $\mathcal{J}$ and $\mathcal{S}_m$ are continuous, bounded from below (by zero) and coercive functions. The proof that $\mathcal{S}_m$ is δ-strongly convex with δ given in (74) is omitted since it can be derived in a very similar way as proof of Proposition 1.

The following important result on strongly convex functions is shown in [34, 36].

Lemma 4 *Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be a δ-strongly convex function, and $x^* \in \mathbb{R}^n$ be a minimizer of function $f(x)$. Then, the following inequality holds:*

$$\frac{\delta}{2}\,\|x - x^*\|_2^2 \;\leqslant\; f(x) - f(x^*) \quad \forall x \in \mathbb{R}^n \;. \tag{75}$$

Finally, in the following proposition we give convergence results for the proposed CNC-MM algorithm.

Proposition 9 *Let $\{x^{(k)}\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the CNC-MM approach in Algorithm 1 applied to the solution of nonconvex nonsmooth optimization problems of the form (1) under hypothesis H1)–H2) in (8)–(9). Then, for any initial guess $x^{(0)} \in \mathbb{R}^n$ the following two statements both hold:*

s1) the sequence $\left\{\mathcal{J}(x^{(k)})\right\}_{k=0}^{\infty}$ is monotonically non-increasing and convergent;

s2) the sequence $\{x^{(k)}\}_{k=0}^{\infty}$ *has the property*

$$\sum_{k=0}^{+\infty} \left\| x^{(k+1)} - x^{(k)} \right\|_2^2 \, < \, +\infty \,, \qquad \textit{hence} \quad \lim_{k \longrightarrow \infty} \left\| x^{(k+1)} - x^{(k)} \right\|_2^2 = 0 \,.$$

Proof We recall that, at any iteration $k \geqslant 0$ of the CNC-MM algorithm, the surrogate function $\mathcal{S}_m(x, x^{(k)})$ is a strongly convex tangent majorant of $\mathcal{J}(x)$ at $x^{(k)}$ and the new iterate $x^{(k+1)}$ is the global minimizer of $\mathcal{S}_m(x, x^{(k)})$. Hence, we can write:

$$\mathcal{J}(x^{(k+1)}) \leqslant \mathcal{S}_m(x^{(k+1)}, x^{(k)}) \leqslant \mathcal{S}_m(x^{(k)}, x^{(k)}) = \mathcal{J}(x^{(k)}) \quad \forall\, k \geqslant 0 \,,$$

so that the sequence $\{\mathcal{J}(x^{(k)})\}_{k=0}^{\infty}$ is monotonically non-increasing. Proof of s1) is completed by noting that $\{\mathcal{J}(x^{(k)})\}_{k=0}^{\infty}$ is bounded from below by zero, hence convergent.

Since $\mathcal{S}_m(x, x^{(k)})$ is δ-strongly convex – see Proposition 8 – we use Lemma 4: inequality (75) with $\mathcal{S}_m(x, x^{(k)})$ in place of $f(x)$ and $x^{(k+1)}$ in place of x^* reads as:

$$\frac{\delta}{2} \left\| x - x^{(k+1)} \right\|_2^2 \leqslant \mathcal{S}_m(x, x^{(k)}) - \mathcal{S}_m(x^{(k+1)}, x^{(k)}) \quad \forall x \in \mathbb{R}^n \,, \quad \forall\, k \geqslant 0 \,. \tag{76}$$

Substituting the iterate $x^{(k)}$ for x in (76), we obtain:

$$\begin{aligned} \frac{\delta}{2} \left\| x^{(k)} - x^{(k+1)} \right\|_2^2 &\leqslant \mathcal{S}_m(x^{(k)}, x^{(k)}) - \mathcal{S}_m(x^{(k+1)}, x^{(k)}) && (77) \\ &\leqslant \mathcal{J}(x^{(k)}) - \mathcal{J}(x^{(k+1)}) \quad \forall k \geqslant 0 \,, && (78) \end{aligned}$$

where (78) comes from $\mathcal{S}_m(x^{(k)}, x^{(k)}) = \mathcal{J}(x^{(k)})$ and $\mathcal{S}_m(x^{(k+1)}, x^{(k)}) \geqslant \mathcal{J}(x^{(k+1)})$. Summing the inequalities (77)–(78) over k yields

$$\begin{aligned} \sum_{k=0}^{\infty} \left\| x^{(k+1)} - x^{(k)} \right\|_2^2 &\leqslant \frac{2}{\delta} \sum_{k=0}^{\infty} \Big[\mathcal{J}(x^{(k)}) - \mathcal{J}(x^{(k+1)}) \Big] && (79) \\ = \frac{2}{\delta} \big(\underbrace{\mathcal{J}(x^{(0)}) - \cancel{\mathcal{J}(x^{(1)})}}_{k=0} + \underbrace{\cancel{\mathcal{J}(x^{(1)})} - \cancel{\mathcal{J}(x^{(2)})}}_{k=1} + \ldots \big) &= \frac{2}{\delta} \big(\mathcal{J}(x^{(0)}) - \mathcal{J}^* \big) , && (80) \end{aligned}$$

where $\mathcal{J}^*$ denotes the (finite) limit of the convergent sequence $\big\{\mathcal{J}(x^{(k)})\big\}_{k=0}^{\infty}$. Since $0 < \delta < +\infty$ and the sequence $\big\{\mathcal{J}(x^{(k)})\big\}_{k=0}^{\infty}$ is monotonically non-increasing, then the right-hand side of (80) is a finite non-negative number and the series on the left-hand side of (79) is convergent. Therefore statement s2) is proved.

7 Numerical Examples

In this section we evaluate experimentally the performance of the proposed CNC-MM approach applied to the solution of nonsmooth nonconvex minimization problems of the form (1) both in terms of speed of convergence and in terms of quality of the approached limit points. In particular, we are interested in comparing our CNC-MM proposal with the popular Q-MM and L-MM methods, which represent the more natural competitors.

In the first numerical example we will consider a simple scalar model problem which allows for a complete control over the cost functional form and for a closed-form solution of the minimization steps, such that the attention can be focused on the speed of convergence of the MM outer iterations. In the second example a less trivial bivariate problem will be considered which allows to evaluate / compare also the quality of the solutions obtained by the Q-MM, L-MM and CNC-MM algorithms. Finally, in Example 3 we investigate the benefit of using nonconvex instead of convex variational models for the restoration of a 1D corrupted signal.

7.1 Example 1

We consider the following simple scalar model problem:

$$\min_{x\in\mathbb{R}} \mathcal{J}(x)\,, \qquad \mathcal{J}(x) \,=\, \frac{1}{2}\,(x-b)^2 + \mu\,\phi_{\log}(x;a)\,, \tag{81}$$

with $b \in \mathbb{R}$, $\mu, a \in \mathbb{R}^*_+$. Problem (81) is a scalar instance of the n-dimensional problem (1) with scalar matrices $A, L = 1$ such that, according to definition (7), $\rho(A, L) = 1$. The cost function $\mathcal{J}(x)$ in (81) is nonsmooth at $x = 0$ and, since condition (5) in Proposition 1 is satisfied, can be convex or nonconvex depending on the parameters μ and a. In particular, according to condition (6) in Proposition 1, $\mathcal{J}(x)$ is strictly convex (or simply convex) if $\mu\,a < \rho(A, L) = 1$ (or $\mu\,a \leqslant 1$), whereas it is nonconvex for $\mu\,a > 1$.

We consider the parameters setting: $b = 1$, $a = 4$, $\mu = 1.4$ ($\overline{\tau}$= 5.6), such that $\mu\,a = 5.6 > 1$ and, hence, $\mathcal{J}(x)$ is nonconvex. The function $\mathcal{J}(x)$, depicted in solid red in Figure 3, has the three critical points:

$$x_1^*\!= 0, \quad x_{2,3}^* = \,\mathrm{sign}(b)\frac{1}{a}\left(\frac{a}{2}\,|b| - \frac{1}{2} \mp \sqrt{\left(\frac{a}{2}\,|b| + \frac{1}{2}\right)^2 - a\,\mu}\,\right) = \{\,0.173\ldots, 0.576\ldots\}\,,$$

where the closed-form formulas for $x_{2,3}^*$ derive easily from a suitable adaptation of (64). We notice that x_1^* is a nonsmooth (global) minimizer of $\mathcal{J}(x)$, whereas x_2^* and x_3^* are a smooth (local) maximizer and a smooth (local) minimizer, respectively.

We solve problem (81) by applying the Q-MM, L-MM and CNC-MM algorithms, based on the construction/minimization of the surrogate functions

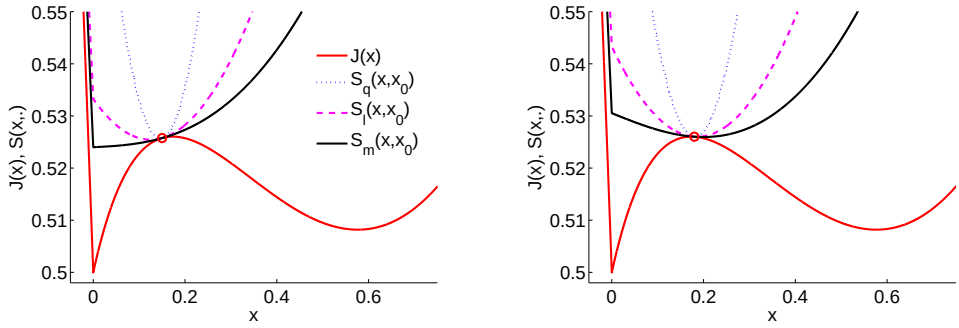

Fig. 3 Surrogate functions $S_q(x, x_0)$, $S_l(x, x_0)$, $S_m(x, x_0)$ defined in (82)–(84), majorizing $\mathcal{J}(x)$ in (81) at the two different initial guesses $x_0 = 0.15$ (left) and $x_0 = 0.18$ (right).

S_q, S_l and S_m in (20), (21) and (47), respectively, which in this specific scalar case read as follows:

$$S_q(x, x_q^{(k)}) = \frac{1}{2}(x-b)^2 + \mu\, w_q^{(k)} \frac{x^2}{2} + C_q^{(k)}\,, \tag{82}$$

$$S_l(x, x_l^{(k)}) = \frac{1}{2}(x-b)^2 + \mu\, w_l^{(k)} |x| + C_l^{(k)}\,, \tag{83}$$

$$S_m(x, x_m^{(k)}) = \frac{1}{2}(x-b)^2 + \underline{\tau}\rho \frac{1}{\hat{a}_m^{(k)}} \phi_{\log}(x; \hat{a}_m^{(k)}) + C_m^{(k)}\,, \tag{84}$$

where the majorization weights $w_q^{(k)}$, $w_l^{(k)}$ in (82)–(83), are given according to the formulas reported in Table 2 for the penalty function $\phi_{\log}$ and where $x_q^{(k)}, x_l^{(k)}, x_m^{(k)}$ denote the kth iterates obtained by the three methods starting from a common initial guess $x_q^{(0)} = x_l^{(0)} = x_m^{(0)} = x_0$. For the CNC-MM approach, we used a value $\underline{\tau} = 0.99$.

In this simple scalar example, the global minimizers of the strictly convex surrogates $S_q(x, x_q^{(k)})$, $S_l(x, x_l^{(k)})$ and $S_m(x, x_m^{(k)})$ defined in (82)–(84) can be computed by means of simple closed-form formulas.

We applied the three considered approaches starting from two different initial guesses $x_0 = 0.15$ and $x_0 = 0.18$, since these two choices yield convergence of the methods toward the two different local minimizers x_1^* and x_3^* of the function $\mathcal{J}$.

In Figure 3 we show the surrogate functions $S_q(x, x_0)$, $S_l(x, x_0)$, $S_m(x, x_0)$ constructed and then minimized by the three approaches at their first iteration for the two initial guesses $x_0 = 0.15$ (on the left) and $x_0 = 0.18$ (on the right). We notice how in both cases the proposed CNC majorant S_m better fits the nonconvex nonsmooth objective function $\mathcal{J}$; this, in principle, holds the potential for a faster convergence towards the local or global minimizers.

In Figure 4 we show the iterates $\big(x_q^{(k)}, \mathcal{J}(x_q^{(k)})\big)$, $\big(x_l^{(k)}, \mathcal{J}(x_l^{(k)})\big)$, $\big(x_m^{(k)}, \mathcal{J}(x_m^{(k)})\big)$ obtained by the three approaches starting from the two initial guesses $x_0 = 0.15$ (top row) and $x_0 = 0.18$ (bottom row). As expected, the different initial

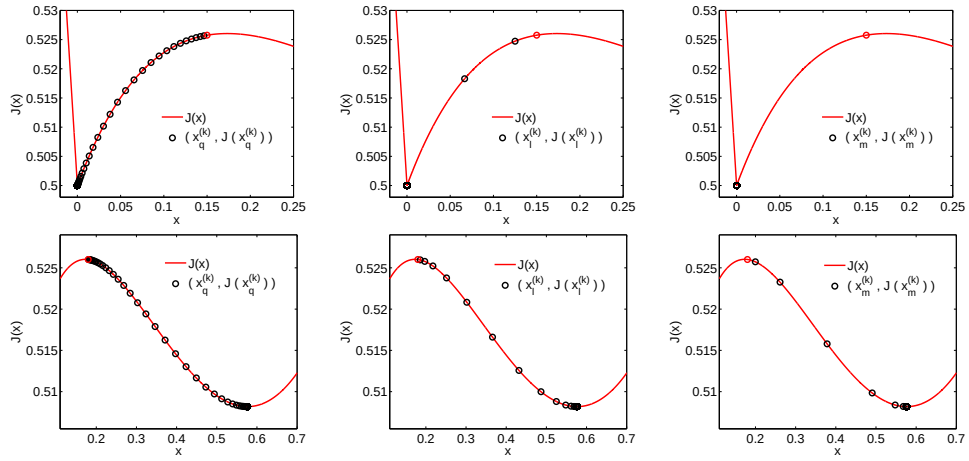

Fig. 4 Iterates $\big(x_q^{(k)}, \mathcal{J}(x_q^{(k)})\big)$, $\big(x_l^{(k)}, \mathcal{J}(x_l^{(k)})\big)$, $\big(x_m^{(k)}, \mathcal{J}(x_m^{(k)})\big)$ obtained by applying MM to the minimization of the function $\mathcal{J}(x)$ in (81) with quadratic, piecewise linear and CNC majorization, starting from the initial guesses $x_0 = 0.15$ (top row) and $x_0 = 0.18$ (bottom row).

guesses lead to different minimizers, but we notice that none of all the three surrogate majorants used in the MM algorithm has the ability to avoid local minima. However, as expected, by using CNC majorization less iterations are required to reach the minimums.

To evaluate quantitatively the speed of convergence of the considered methods, we define the two error sequences:

$$e_x^{(k)} := \big|x^{(k)} - x^*\big|, \qquad e_{\mathcal{J}}^{(k)} := \big|\mathcal{J}^{(k)} - \mathcal{J}^*\big|, \qquad k = 0, 1, 2 \ldots, \tag{85}$$

where $x^{(k)}$ and $\mathcal{J}^{(k)} := \mathcal{J}\big(x^{(k)}\big)$ denote the iterates and the associated function values generated by the algorithms, x^* and $\mathcal{J}^* := \mathcal{J}\big(x^*\big)$ are the limits of the two sequences $x^{(k)}$ and $\mathcal{J}^{(k)}$, which are known thanks to the available closed-form expressions for the minimizers x_1^* and x_3^* given above. In Figure 5 we show the sequences of error quotients $e_x^{(k+1)}/e_x^{(k)}$ and $e_{\mathcal{J}}^{(k+1)}/e_{\mathcal{J}}^{(k)}$ obtained by the three algorithms Q-MM, L-MM and CNC-MM when starting from the two initial guesses $x_0 = 0.15$ (left column) and $x_0 = 0.18$ (right column), and thus converging towards the minimum points $\big(x_1^*, \mathcal{J}\big(x_1^*\big)\big)$ and $\big(x_3^*, \mathcal{J}\big(x_3^*\big)\big)$. For all the reported plots we stopped the iterations as soon as the error sequences in (85) drop below a prescribed threshold equal to 10^{-13}. From the plots shown in the right-most column of Figure 5, we notice that all the methods converge linearly towards the smooth (local) minimum. However, CNC-MM exhibits a smaller convergence factor than L-MM and Q-MM, so that fewer iterations are required to achieve the prescribed error bounds. For what concerns the nonsmooth (global) minimum, the plots shown in the left-most column of Figure 5 suggest that Q-MM again converges linearly, whereas L-MM and CNC-MM converge super-linearly. Actually, in this scalar example

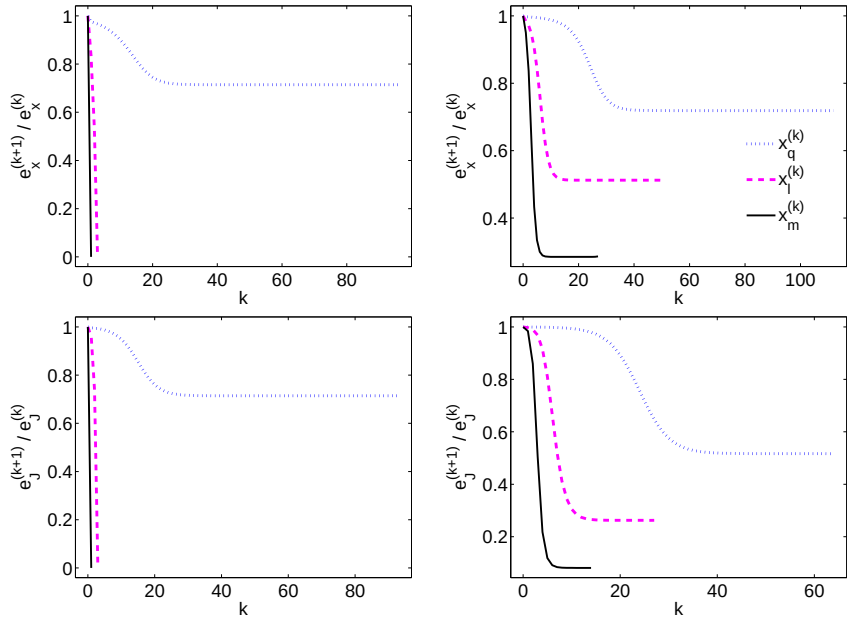

Fig. 5 Empirical analysis of the asymptotic convergence of Q-MM, L-MM and CNC-MM approaches applied to the solution of problem (81) starting from the initial guesses $x_0 = 0.15$ (left column) and $x_0 = 0.18$ (right column).

the L-MM and CNC-MM methods achieve the nonsmooth minimizer $x_1^* = 0$ (up to machine precision) in a finite number of iterations, namely 3 iterations for L-MM and a unique iteration for CNC-MM.

7.2 Example 2

This example illustrates the minimization of a nonconvex objective function $\mathcal{J}\colon \mathbb{R}^2 \to \mathbb{R}$. The minimization problem is

$$\min_{x\in\mathbb{R}^2} \mathcal{J}(x), \qquad \mathcal{J}(x) \;=\; \frac{1}{2}\,\|Ax-b\|_2^2 + 4\phi_{\log}(x_1;a) + 4\phi_{\log}(x_2;a)\,, \tag{86}$$

where $x = (x_1, x_2)^T$, $b = (2.5, 1)^T$, $A = (2, 2; 1, -1)$, and the concavity parameter is $a = 2$. We identify $P(x) = 4\phi(x_1; a) + 4\phi(x_2; a)$ as the penalty function (which is nonconvex). For this value of a, the objective function $\mathcal{J}$ is also nonconvex. Figure 6 illustrates $\mathcal{J}$ in the positive quadrant. The function has a local minimum at $x = (0, 0.622...)$ and a global minimum at $x = (0.679..., 0)$.

We investigate the minimization of $\mathcal{J}$ using the iterative MM process, which approximates $\mathcal{J}$ at each iteration by a surrogate function. In this example, we compare the MM process using the piecewise linear surrogate function S_l and the proposed CNC surrogate function S_m. The surrogate function S_l is obtained by majorizing the penalty function P by the ℓ_1 norm. On the other

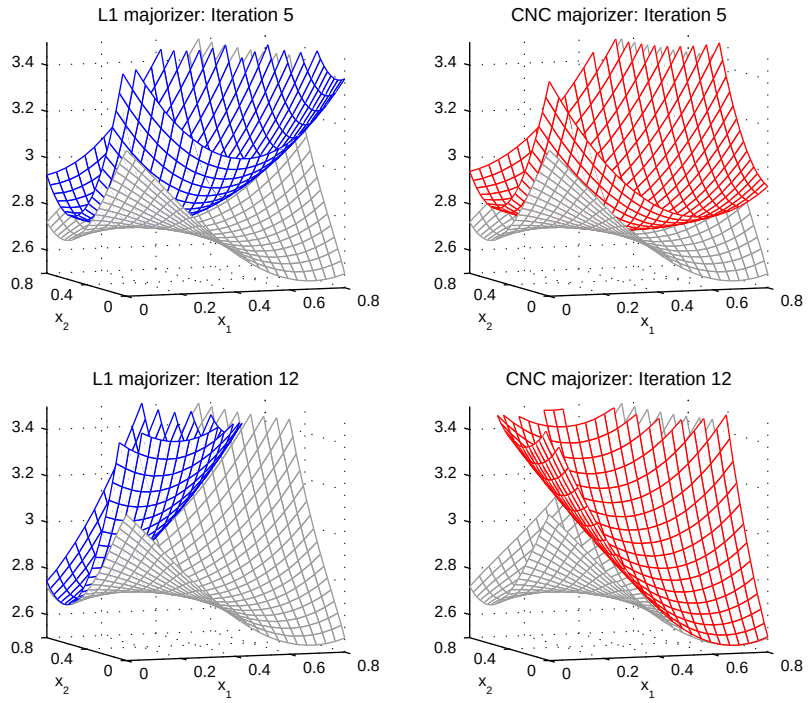

Fig. 6 Example 2. Iterations 5 and 12 of the L-MM and CNC-MM algorithms. In both case, the nonconvex objective function $\mathcal{J}$ (in gray) is majorized by a convex surrogate function.

hand, the surrogate function S_m is obtained by majorizing P by a nonconvex function. In both cases, the surrogate function is always convex. We initialize the MM process in both cases with the starting point $x^{(0)} = (0.35, 0.7)$.

Figure 6 illustrates the surrogate functions S_l and S_m at iterations 5 and 12. The CNC surrogate function S_m more closely approximates the objective function $\mathcal{J}$; hence, it yields a different sequence of iterates $x^{(k)}$. As shown in Fig. 6, the MM process using ℓ_1 norm majorization leads to a local minimum. However, MM process using the proposed CNC majorization method leads to the global minimum. For each case, the history of the iterates is shown in Fig. 7.

We note that the convergence of the CNC approach to the global minimum does depend on the starting point $x^{(0)}$. For starting points that are too close to a local minimum, the CNC approach will converge to it as will the ℓ_1 norm majorization method. However, this examples demonstrates that, due to its better approximation of the objective function, the CNC approach is more likely to avoid local minima.

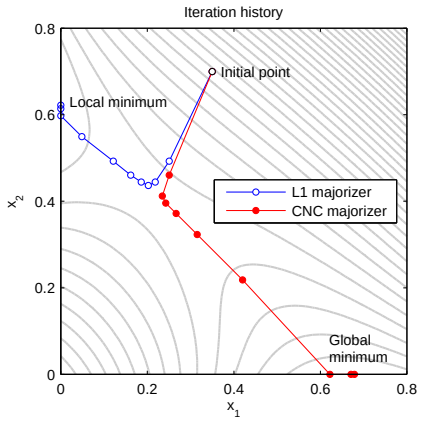

Fig. 7 Example 2. History of iterates.

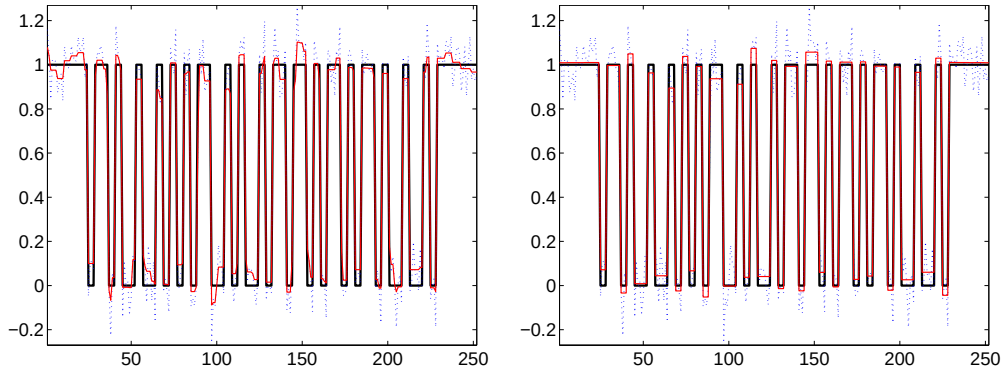

Fig. 8 Example 3. Original uncorrupted piecewise constant signal (solid black line), noise-corrupted signal (dotted blue line), restored signals (solid red line) by model (87) with $\bar{\tau} = 0$, i.e. by ROF (left: ISNR = 3.30) and with $\bar{\tau} = 50$ (right: ISNR=7.61).

7.3 Example 3

In this last example we consider a more realistic, higher dimensional application of our method, namely the restoration of piecewise constant 1D signals corrupted by additive white Gaussian noise. The restored signals are obtained as the minimizers of the following variational model:

$$\min_{x\in\mathbb{R}^n} \mathcal{J}(x)\,, \qquad \mathcal{J}(x) \;=\; \frac{1}{2}\,\|x-b\|_2^2 \;+\; \mu \sum_{i=1}^{n-1} \phi_{\log}\big((Lx)_i; a\big)\,, \tag{87}$$

with L representing the (forward) finite difference approximation of the first-order derivative operator with Neumann boundary conditions. Hence, we have $\sigma_{A,\min} = 1$, $\sigma_{L,\max} = 2$ and, according to (7), $\rho = 1/4$. We notice that model (87) represents a generalization of the popular ROF model for signal denoising

$\bar{\tau}$	0	0.5	0.99	2.0	25	50	100	200	500	5000
$\mu \cdot 10^3$	80	120	147	184	456	529	589	631	664	688
ISNR	3.30	4.59	5.08	5.59	7.40	7.61	7.73	7.79	7.81	7.82

Table 3 Example 3. Quality of the restored signals (ISNR) obtained by using the variational model in (87) with increasing values of the parameter $\bar{\tau}$.

[45]. In fact, according to property (3) of the penalty function $\phi(\,\cdot\,;a)$, (87) coincides with the ROF model when $a = 0$. Moreover, according to (9), for a fixed μ, when a increases $\bar{\tau}$ increases, and the model passes from convex ($0 < \bar{\tau} \leqslant 1$) to nonconvex ($\bar{\tau} > 1$).

We consider the restoration of a piecewise constant binary (0/1) signal of dimension $n = 252$ representing a bar-code (of type USS-39) that has been corrupted by additive zero-mean white Gaussian noise with standard deviation $\sigma = 0.078$. Figure 8 shows the noise-free signal (solid line) and the degraded signal (dotted-blue line).

The quality of the restored signals is evaluated by the Signal-to-Noise Ratio (SNR) defined as $\mathrm{SNR}(x^*, \bar{x}) := 10 \log_{10} \big(\|\bar{x} - E[\bar{x}]\|_2^2 \;/\; \|x^* - \bar{x}\|_2^2 \big)$, where $x^* \in \mathbb{R}^n$ is the computed estimate of the uncorrupted signal $\bar{x} \in \mathbb{R}^n$ and $E[\bar{x}]$ denotes the mean value of $\bar{x}$. More precisely, the Improved Signal-to-Noise Ratio (ISNR), defined as $\mathrm{ISNR}(x^*, \bar{x}, b) := \mathrm{SNR}(x^*, \bar{x}) - \mathrm{SNR}(b, \bar{x})$, provides a quantitative measure of the improvement in the quality of the denoised image: a high ISNR value indicates that x^* is an accurate approximation of $\bar{x}$.

For both the ROF and CNC-MM algorithms, we used as initial iterate $x^{(0)} \in \mathbb{R}^n$ the constant signal with value equal to the mean of the observed noisy signal b and the iterations are stopped as soon as the two successive iterates satisfies

$$\|x^{(k)} - x^{(k-1)}\|_2 \;/\; \|x^{(k-1)}\|_2 \;<\; \epsilon\,, \tag{88}$$

where $\epsilon > 0$ is a user-specified threshold. We used $\epsilon = 10^{-5}$.

We used a parameter value $\underline{\tau} = 0.99$ for the CNC-MM majorization step, such that at any iteration the constructed CNC surrogate is (almost) maximally tight to the objective functional $\mathcal{J}(x)$ in (87). For what concerns the CNC-MM minimization step by ADMM, we used the parameter values $\beta = 10$. We notice that the chosen value of the penalty parameter β satisfies condition (63) for convexity of the ADMM subproblems in (57). In particular, experiments demonstrated that faster convergence of the ADMM-based minimization step is always obtained for β in the range $[5, 40]$, such that constraint (63) does not limit the efficiency of the CNC-MM algorithm.

In Table 3 we report the ISNR values associated with the obtained restored signals for different values of the parameter $\bar{\tau}$, ranging from the convex case ($\bar{\tau} = 0$) to the CNC ($0 < \bar{\tau} \leqslant 1$) and the nonconvex ($\bar{\tau} > 1$) cases. We notice that, according to the definition of $\bar{\tau}$, there exists an infinite number of pairs $(\mu, a = \bar{\tau}\rho/\mu)$ yielding a given $\bar{\tau}$. Hence, in Table 3 for each considered $\bar{\tau}$ we report the highest ISNR value achieved by letting μ (and, accordingly, a) vary and the associated μ.

We remark that beyond the greatest considered value $\bar{\tau} = 5000$ the ISNR improvement is negligible. Results in Table 3 strongly indicate that higher quality restorations can be achieved by pushing model (87) beyond its convexity limits. This confirms the usefulness of the proposed CNC-MM minimization approach.

In Figure 8 we show the restoration results obtained for the cases $\bar{\tau} = 0$, i.e. ROF (left), and $\bar{\tau} = 50$ (right) of Table 3. It is well visible how the restored signal obtained by using the nonconvex model minimized by the proposed CNC-MM algorithm is a much better approximation of the original piecewise constant signal.

In the above nonconvex regime, the efficacy of the proposed MM procedure with CNC surrogate $S_m(x, x^{(k)})$ (CNC-MM Algorithm 1) is then finally compared with the MM procedure using $S_l(x, x^{(k)})$ (L-MM) and $S_q(x, x^{(k)})$ (Q-MM) surrogates. In particular, the ISNR values and the number of outer/inner iterations for a fixed $\mu = 0.529$ are: ISNR=7.43 for L-MM, with 94/3663 outer/inner iterations, ISNR=7.44 for Q-MM, with 78/16758 outer/inner iterations, ISNR=7.61 for CNC-MM, with 8/519 outer/inner iterations.

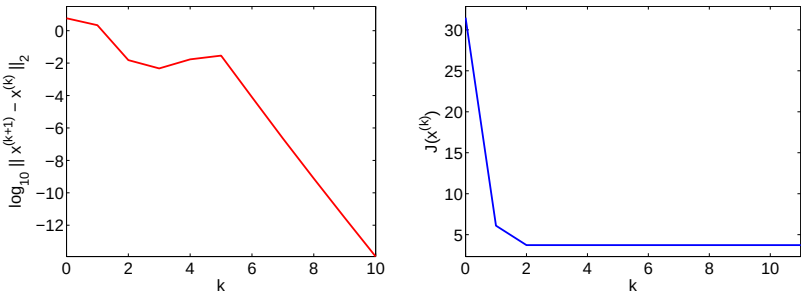

Fig. 9 Example 3. Plots providing empirical evidence of numerical convergence of the proposed CNC-MM minimization algorithm for the case $\bar{\tau} = 50$ in Table 3.

We conclude this example by presenting a short empirical investigation on numerical convergence of the proposed CNC-MM minimization algorithm. In Figure 9 we report two convergence plots obtained by applying the CNC-MM method to the (nonconvex) case $\bar{\tau} = 50$ in Table 3. In particular, to better highlight the convergence behavior, both the outer and the inner (ADMM) iterations of the CNC-MM algorithm have been stopped according to (88) with a very small tolerance $\epsilon = 10^{-15}$. The plots in Figure 9 provide strong evidence in favor of convergence of the generated iterates sequence as well as of monotonicity of the sequence of associated function values.

8 Conclusions

We proposed a novel MM strategy for the solution of a certain class of nonsmooth nonconvex optimization problems: the objective function is the sum

of a strongly convex quadratic (fidelity) term and a nonconvex nonsmooth sparsity-promoting regularization term. Majorizing the nonconvex regularizer with a nonconvex surrogate function, designed so that the total surrogate function is strongly convex, allows for a tight approximation of the objective function and, hence, for fast convergence and robustness to local minimizers. A suitable ADMM-based algorithm has been presented for the efficient solution of the minimization step. In particular, the solution of the linear system, which represents the most computationally expensive step of Algorithm 1, is carried out in an efficient way due to the fact that the coefficient matrix does not change during inner/outer iterations. A preliminary convergence analysis for the CNC-MM proposal has been provided. Numerical experiments demonstrate the effectiveness of the proposed approach, when compared with the IRL2 and IRL1 algorithms.

Acknowledgements

We would like to thank the referees for comments that lead to improvements of the presentation. Research by IS was supported by the NSF (USA) under Grant No. CCF-1525398. Research by SM and FS was supported in part by the National Group for Scientific Computation (GNCS-INDAM), Research Projects 2015.

Appendix

Proof of Proposition 3.3.

Proof By applying formula (4) in Lemma 2, namely $\phi(t;a) = u(t;a)+|t|$, after simple algebraic manipulations the functional $\mathcal{J}(x)$ in (1) can be equivalently rewritten as follows:

$$\mathcal{J}(x) \;=\; \underbrace{\frac{1}{2}\,\|Ax\|_2^2 + \sum_{i=1}^{s} \mu_i u\left((Lx)_i; a_i\right)}_{\mathcal{J}_1(x)} \;+\; \underbrace{\frac{1}{2}\,\|b\|_2^2 - b^T Ax \;+\; \sum_{i=1}^{s} \mu_i |(Lx)_i|}_{\mathcal{J}_2(x)} \;. \tag{89}$$

Since functional $\mathcal{J}_2(x)$ in (89) is convex, convexity of $\mathcal{J}(x)$ follows from convexity of $\mathcal{J}_1(x)$, which is twice continuously differentiable due to statement 1) of Lemma 2. Hence, a sufficient condition for $\mathcal{J}(x)$ being strictly convex is that the Hessian matrix $H(x)$ of functional $\mathcal{J}_1(x)$ in (89) is positive definite for all $x \in \mathbb{R}^n$, that is:

$$H(x) \;=\; \underbrace{A^T A}_{H_A} \;-\; \underbrace{L^T\, \Gamma(x)\, L}_{H_L(x)} \;\succ\; 0 \quad \forall x \in \mathbb{R}^n \;, \tag{90}$$

where $\Gamma(x)$ is the $s \times s$ diagonal matrix depending on x defined as:

$$\Gamma(x) = \mathrm{diag}\big(\gamma_1(x), \ldots, \gamma_s(x)\big) \;, \qquad \gamma_i(x) = \; -\,\mu_i\, u''((Lx)_i; a_i)\,. \tag{91}$$

Since $\mu_i > 0$ by assumption, from statement 2) of Lemma 2 it follows that:

$$\gamma_i(x) \in [\,0\,,\,\mu_i a_i\,] \quad \forall\, x \in \mathbb{R}^n, \quad \forall\, i \in \{1,\ldots,s\}\,. \tag{92}$$

Hence, the two $n \times n$ matrices H_A and $H_L(x)$ in (90) are both at least positive semi-definite, if not positive definite, for any $x \in \mathbb{R}^n$. We notice that if matrix H_A is only positive semi-definite, that is $\ker\{A^T A\} \neq \{0\}$, there is no possibility for the Hessian matrix $H(x)$ in (90) to be positive definite for all $x \in \mathbb{R}^n$. This justifies condition (5) for strict convexity of $\mathcal{J}(x)$.

To further investigate positive-definiteness of matrix $H(x)$ in (90), we introduce the Singular Value Decomposition (SVD) of matrices $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{s\times n}$:

$$\begin{aligned} A &= U_A\,\Sigma_A V_A^T,\; U_A \in \mathbb{R}^{m\times n}, \Sigma_A \in \mathbb{R}^{n\times n}, V_A \in \mathbb{R}^{n\times n}, \\ L &= U_L\,\Sigma_L\,V_L^T,\; U_L \in \mathbb{R}^{s\times p}, \Sigma_L \in \mathbb{R}^{p\times p}, V_L \in \mathbb{R}^{n\times p}, \quad p := \min\{s,n\}, \end{aligned} \tag{93}$$

where in the SVD of matrix $A \in \mathbb{R}^{m\times n}$ we are implicitly assuming that $m \geqslant n$, since otherwise condition (5) can not be satisfied. We recall that Σ_A and Σ_L in (93) are diagonal matrices containing the singular values of matrices A and L, respectively, while U_A, V_A and U_L, V_L are orthogonal matrices containing the left and right singular vectors of matrices A and L, respectively, and are such that $U_A^T U_A = V_A^T V_A = V_A V_A^T = I_n$ and $U_L^T U_L = V_L^T V_L = I_p$.

By substituting (93) into the positive-definiteness condition (90), we obtain:

$$H(x) \;=\; \underbrace{V_A \Sigma_A^2 V_A^T}_{H_A} \;-\; \underbrace{V_L \Sigma_L U_L^T\, \Gamma(x)\, U_L \Sigma_L V_L^T}_{H_L(x)} \;\succ\; 0 \quad \forall x \in \mathbb{R}^n\,. \tag{94}$$

In order to obtain sufficient conditions for (94) being satisfied, we introduce a lower bound (in terms of positive-definiteness) $\underline{H}_A$ for H_A:

$$\underline{H}_A := V_A\, \underline{\Sigma}_A^2\, V_A^T \;=\; V_A\, \sigma_{A,\min}^2 I_n\, V_A^T \;=\; \sigma_{A,\min}^2 I_n \;\preccurlyeq\; H_A\,, \tag{95}$$

and an upper bound $\overline{H}_L$ for $H_L(x)$:

$$\begin{aligned} \overline{H}_L \;&:=\; V_L\, \overline{\Sigma}_L U_L^T\; \overline{\Gamma}\, U_L\, \overline{\Sigma}_L V_L^T \\ &=\; V_L\, \sigma_{L,\max} I_p\, U_L^T\, \mathrm{diag}(\mu_1 a_1, \ldots, \mu_s a_s)\, U_L\, \sigma_{L,\max} I_p\, V_L^T \\ &=\; \sigma_{L,\max}^2\, V_L U_L^T\, \mathrm{diag}(\mu_1 a_1, \ldots, \mu_s a_s)\, U_L V_L^T \;\succcurlyeq\; H_L(x) \quad \forall x \in \mathbb{R}^n \end{aligned} \tag{96}$$

where $\sigma_{A,\min}$ and $\sigma_{L,\max}$ denote the minimum and maximum among the singular values of matrices A and L, respectively, and where the upper bound $\overline{\Gamma}$ comes from properties (92) of the diagonal matrix $\Gamma(x)$ defined in (91). By substituting the lower bound $\underline{H}_A$ in (95) for H_A and the upper bound $\overline{H}_L$ in (96) for $H_L(x)$ into the definition of matrix $H(x)$ in (94), and introducing the matrix

$$X := U_L V_L^T \in \mathbb{R}^{s\times n}\,, \tag{97}$$

we obtain a lower bound $\underline{H}$ for $H(x)$:

$$\begin{aligned}\underline{H} &:= \underline{H}_A - \overline{H}_L \\ &= \sigma^2_{A,\min} I_n - \sigma^2_{L,\max} X^T \mathrm{diag}(\mu_1 a_1, \ldots, \mu_s a_s)\, X \preccurlyeq H(x) \quad \forall\, x \in \mathbb{R}^n \end{aligned} \tag{98}$$

We notice that, since U_L and V_L are orthogonal matrices, the matrix $X \in \mathbb{R}^{s\times n}$ defined in (97) has full (column and/or row) rank, that is $\mathrm{rank}\{X\} = \min\{s, n\} = p$. To conclude the proof, we consider separately the two cases $s \geqslant n$ (square or tall matrix) and $s < n$ (wide matrix) for the linear operator $L \in \mathbb{R}^{s\times n}$.

Case $s \geqslant n$. In this case, since $p := \min\{s, n\} = n$, the SVD in (93) of the square or tall matrix $L \in \mathbb{R}^{s\times n}$ reads as follows: $L = U_L\, \Sigma_L\, V_L^T$, $U_L \in \mathbb{R}^{s\times n}$, $\Sigma_L \in \mathbb{R}^{n\times n}$, $V_L \in \mathbb{R}^{n\times n}$, where the orthogonal matrices U_L and V_L are such that $U_L^T U_L = V_L^T V_L = V_L V_L^T = I_n$, and the matrix $X \in \mathbb{R}^{s\times n}$ defined in (97) satisfies:

$$X^T X = V_L U_L^T U_L V_L^T = I_n \,, \tag{99}$$

that is the n ($\leqslant s$) columns of X are s-dimensional orthonormal vectors. By substituting the expression (99) for I_n in (98), the lower bound matrix $\underline{H}$ can be equivalently rewritten as follows:

$$\begin{aligned}\underline{H} &= \sigma^2_{A,\min} X^T X - \sigma^2_{L,\max} X^T \mathrm{diag}(\mu_1 a_1, \ldots, \mu_s a_s)\, X \\ &= X^T \mathrm{diag}\big(\, \sigma^2_{A,\min} - \sigma^2_{L,\max}\, \mu_1 a_1,\ \ldots\ , \sigma^2_{A,\min} - \sigma^2_{L,\max}\, \mu_s a_s \big)\, X \end{aligned} \tag{100}$$

Recalling that $\underline{H}$ in (100) is a lower bound of $H(x)$ in (90) for any $x \in \mathbb{R}^n$ and that the matrix $X \in \mathbb{R}^{s\times n}$ in (100) has full column rank, it follows from Lemma 1 that:

$$H(x) \succcurlyeq \underline{H} \succ 0 \ \ \forall x \in \mathbb{R}^n \quad \text{if} \quad \sigma^2_{A,\min} - \sigma^2_{L,\max}\, \mu_i a_i > 0 \ \ \forall\, i \in \{1, \ldots, s\}\,, \tag{101}$$

thus proving the second condition for strict convexity of $\mathcal{J}(x)$ in (6).

Case $s < n$. In this case, $p := \min\{s, n\} = s$ and the SVD in (93) of the wide matrix $L \in \mathbb{R}^{s\times n}$ is $L = U_L\, \Sigma_L\, V_L^T$, $U_L \in \mathbb{R}^{s\times s}$, $\Sigma_L \in \mathbb{R}^{s\times s}$, $V_L \in \mathbb{R}^{n\times s}$, where the orthogonal matrices U_L and V_L are such that $U_L^T U_L = U_L U_L^T = V_L^T V_L = I_s$, and the matrix $X \in \mathbb{R}^{s\times n}$ defined in (97) satisfies: $X X^T = U_L V_L^T V_L U_L^T = I_s$, that is the s ($< n$) rows of X are n-dimensional orthonormal vectors. Hence, it is always possible to build a square orthogonal matrix $\widetilde{X} \in \mathbb{R}^{n\times n}$ defined as follows:

$$\widetilde{X} := \big(\, X\ \tilde{v}_1\ \ldots\ \tilde{v}_{n-s} \,\big)^T, \tilde{v}_i \in \mathbb{R}^n \ \text{ such that } \ \widetilde{X}\widetilde{X}^T = \widetilde{X}^T\widetilde{X} = I_n \,. \tag{102}$$

Based on (102), the lower bound matrix $\underline{H}$ defined in (98) is rewritten as follows:

$$\begin{aligned}\underline{H} &= \sigma^2_{A,\min} \widetilde{X}^T\widetilde{X} - \sigma^2_{L,\max}\, \widetilde{X}^T \mathrm{diag}(\mu_1 a_1, \ldots, \mu_s a_s, \underbrace{0, \ldots, 0}_{n-s\ entries})\, \widetilde{X} \\ &= \widetilde{X}^T \mathrm{diag}\big(\, \sigma^2_{A,\min} - \sigma^2_{L,\max}\, \mu_1 a_1\,,\ \ldots\ , \sigma^2_{A,\min} - \sigma^2_{L,\max}\, \mu_s a_s\,, \\ &\qquad \underbrace{\sigma^2_{A,\min}\,,\ \ldots\ , \sigma^2_{A,\min}}_{n-s\ entries} \big)\, \widetilde{X}\,. \end{aligned} \tag{103}$$

Since $\underline{H}$ in (103) is a lower bound of $H(x)$ in (90) for any $x \in \mathbb{R}^n$ and $\widetilde{X} \in \mathbb{R}^{n\times n}$ in (103) has full (column and row) rank, from Lemma 1 it follows that:

$$H(x) \succcurlyeq \underline{H} \succ 0 \ \forall x \in \mathbb{R}^n \quad \text{if} \quad \begin{cases} \sigma^2_{A,\min} - \sigma^2_{L,\max}\,\mu_i a_i > 0 \quad \forall\, i \in \{1,\ldots,s\} \\ \sigma^2_{A,\min} > 0 \end{cases} . \tag{104}$$

Since $\ker\{A^T A\} = \{0\}$, the second condition $\sigma^2_{A,\min} > 0$ in (104) is always satisfied, while the first condition is equivalent to the convexity condition in (6).

Proof of Proposition 3.

Proof By substituting v for t in (12), and c_m given in (13), we obtain (14).

Recalling the definition of w_m in (13), the first-order partial derivative m_t of the majorant function m in (12) with respect to t, for $t, v \in \mathbb{R} \setminus \{0\}$, is given by:

$$m_t(t, v; a_m) = \frac{\phi'(v; a)}{\phi'(v; a_m)}\ \phi'(t; a_m)\,. \tag{105}$$

Substituting v for t in (105) we have (15). In the case $v = 0$, the majorant function in (12) reduces to $m(t, 0; a_m) = \phi(t; a_m)$. Since both the majorized and the majorant functions belong to the family of penalty functions ϕ defined in Section 2, it follows from assumption *A6*) that $m_t(0^{\pm}, 0; a_m) = \phi'(0^{\pm}; a) = \pm 1$, hence (16).

Since both the majorized function $\phi(t; a)$ and the majorizing function $m(t, v; a_m)$ are continuous and even in t for any $v \in \mathbb{R}$ due to assumptions A1) and A2), it is sufficient to prove (17) for $v > 0$ and $t > 0$. Noting that $\phi(t; a)$ and $m(t, v; a_m)$ are both continuously differentiable in t for $t > 0$ thanks to assumption A3), we have:

$$m(t, v; a_m) = m(v, v; a_m) + \int_v^t m_t(\xi, v; a_m)\, d\xi\,, \tag{106}$$

$$\phi(t; a) = \phi(v; a) + \int_v^t \phi'(\xi; a)\, d\xi\,, \tag{107}$$

Hence, by subtracting (107) from (106) and recalling (14), we obtain:

$$m(t, v; a_m) - \phi(t; a) = \int_v^t \big(m_t(\xi, v; a_m) - \phi'(\xi; a)\big)\, d\xi\,. \tag{108}$$

Given the definition of m_t in (105), we can thus write:

$$\begin{aligned} m(t,v;a_m) - \phi(t;a) &= \int_v^t \left(\frac{\phi'(v;a)}{\phi'(v;a_m)}\, \phi'(\xi;a_m) - \phi'(\xi;a) \right) d\xi \\ &= \int_v^t \left(\frac{\phi'(v;a)}{\phi'(v;a_m)} - \frac{\phi'(\xi;a)}{\phi'(\xi;a_m)} \right) \phi'(\xi;a_m)\; d\xi \\ &= \int_v^t \big(h(v) - h(\xi)\big)\; \phi'(\xi;a_m)\, d\xi \\ &= \int_v^t h'(\vartheta)(v-\xi)\; \phi'(\xi;a_m)\; d\xi\,, \end{aligned} \tag{109}$$

where in the third equality we introduced the function $h : \mathbb{R}_+^* \to \mathbb{R}_+^*$ defined as:

$$h(z) \;=\; \frac{\phi'(z;a)}{\phi'(z;a_m)}\,, \quad z>0\,, \tag{110}$$

and in the last equality (109), which is valid for some ϑ between v and ξ, we replaced the first-order Taylor's expansion of h around ξ. The first-order derivative of function h in (110) is guaranteed to exist for any $z > 0$ due to assumption A3) and is as follows:

$$\begin{aligned} h'(z) &= \frac{\phi''(z;a)\phi'(z;a_m) - \phi'(z;a)\phi''(z;a_m)}{\left(\phi'(z;a_m)\right)^2} \\ &= \frac{\phi'(z;a)}{\phi'(z;a_m)} \left(\frac{\phi''(z;a)}{\phi'(z;a)} - \frac{\phi''(z;a_m)}{\phi'(z;a_m)} \right) \\ &\leqslant 0 \quad \forall\, z>0,\; 0<a_m<a\,, \end{aligned} \tag{111}$$

where the last inequality (111) follows from assumption A4) and assumption A7) with $a_1 = a_m$, $a_2 = a$.

We finally rewrite (109) taking into consideration the integration extremes:

$$m(t,v;a_m) - \phi(t;a) \;=\; \begin{cases} \displaystyle\int_v^t (\xi - v)\, (-h'(\vartheta))\; \phi'(\xi;a_m)\; d\xi & \text{if } \; t > v \\ 0 & \text{if } \; t = v\;. \\ \displaystyle\int_t^v (v - \xi)\, (-h'(\vartheta))\; \phi'(\xi;a_m)\; d\xi & \text{if } \; t < v \end{cases} \tag{112}$$

Recalling (111) and assumption A4), we can conclude that the two integrand functions in (112) are both non negative for any ξ in their associated integration domain, for any possible integration domain defined by $t, v > 0$, for any $0 < a_m < a$, hence (17).

References

1. M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, 19(9):2345–2356, September 2010.
2. M. S. Asif and J. Romberg. Fast and accurate algorithms for re-weighted l_1-norm minimization. *IEEE Trans. Signal Process.*, 61(23):5905–5916, December 2013.
3. J. Baglama and L. Reichel. Augmented implicitly restarted lanczos bidiagonalization methods. *SIAM Journal on Scientific Computing*, 27(1):19–42, 2005.
4. H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.
5. A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, 2(1):183–202, 2009.
6. A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.
7. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
8. E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
9. E. J. Candès, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, 2008.
10. P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, 6(2):298–311, 1997.
11. R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *IEEE Int. Symp. Biomed. Imag. (ISBI)*, pages 262–265, 2009.
12. P.-Y. Chen and I. W. Selesnick. Group-sparse signal denoising: Non-convex regularization, convex optimization. *IEEE Trans. Signal Process.*, 62(13):3464–3478, July 2014.
13. E. Chouzenoux, A. Jezierska, J. Pesquet, and H. Talbot. A majorize-minimize subspace approach for $\ell_2 - \ell_0$ image regularization. *SIAM J. Imag. Sci.*, 6(1):563–591, 2013.
14. E. Chouzenoux, J.-C. Pesquet, and A. Repetti. Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function. *J. Optim. Theory Appl.*, 162(1):107–132, 2014.
15. P. L. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optim.*, 18(4):1351–1376, 2008.
16. P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke et al., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, 2011.
17. I. Daubechies, R. DeVore, M. Fornasier, and C. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Comm. Pure App. Math.*, 63(1):1–38, January 2010.
18. Y. Ding and I. W. Selesnick. Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization. *IEEE Signal Processing Letters*, 22(9):1364–1368, September 2015.
19. T.-M.-T. Do and T. Artières. Regularized bundle methods for convex and non-convex risks. *The Journal of Machine Learning Research*, 13(1):3539–3583, 2012.
20. D.L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, April 2006.
21. M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. Image Process.*, 16(12):2980–2991, 2007.
22. M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, 12(8):906–916, 2003.
23. D. Geman and Y. Chengda. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.*, 4(7):932–946, 1995.
24. D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. and Machine Intel.*, 14(3):367–383, March 1992.

25. T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imag. Sci.*, 2(2):323–343, 2009.
26. G. Huang, A. Lanza, S. Morigi, L.Reichel, and F. Sgallari. A majorization-minimization generalized krylov subspace method for $\ell_p - \ell_q$ image restoration. *submitted*, 2016.
27. M. W. Jacobson and J. A. Fessler. Properties of mm algorithms on convex feasible sets: Extended version. *Technical Report, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122*, 353, 2004.
28. K. Lange, E. C. Chi, and H. Zhou. A brief survey of modern optimization for statisticians. *Int. Stat. Rev.*, 82(1):46–70, 2014.
29. A. Lanza, S. Morigi, L. Reichel, and F. Sgallari. A generalized Krylov subspace method for $\ell_p - \ell_q$ minimization. *SIAM Journal on Scientific Computing*, 37(5):S30–S50, 2015.
30. A. Lanza, S. Morigi, and F. Sgallari. Convex image denoising via non-convex regularization. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale Space and Variational Methods in Computer Vision*, volume 9087 of *Lecture Notes in Computer Science*, pages 666–677. Springer, 2015.
31. A. Lanza, S. Morigi, and F. Sgallari. Constrained $TV_p - \ell_2$ model for image restoration. *Journal of Scientific Computing*, 68(1):64–91, 2016.
32. A. Lanza, S. Morigi, and F. Sgallari. Convex image denoising via non-convex regularization with parameter selection. *Journal of Mathematical Imaging and Vision*, pages 1–26, 2016.
33. L. Laporte, R. Flamary, S. Canu, S. Déjean, and J. Mothe. Nonconvex regularizations for feature selection in ranking with sparse SVM. *IEEE Trans. on Neural Networks and Learning Systems*, 25(6):1118–1130, 2014.
34. J. Mairal. Incremental majorization-minimization optimization with application to large-scale machine learning. *SIAM J. Optim.*, 25(2):829–855, 2015.
35. N. Mourad and J. P. Reilly. Minimizing nonconvex functions for sparse vector reconstruction. *IEEE Trans. Signal Process.*, 58(7):3485–3496, 2010.
36. Y. Nesterov et al. Gradient methods for minimizing composite objective function, 2012.
37. M. Nikolova. Estimation of binary images by minimizing convex criteria. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 108–112 vol. 2, 1998.
38. M. Nikolova. Markovian reconstruction using a GNC approach. *IEEE Trans. Image Process.*, 8(9):1204–1220, 1999.
39. M. Nikolova. Energy minimization methods. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, chapter 5, pages 138–186. Springer, 2011.
40. M. Nikolova, M. K. Ng, and C.-P. Tam. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Trans. Image Process.*, 19(12):3073–3088, 2010.
41. A. Parekh and I. W. Selesnick. Convex denoising using non-convex tight frame regularization. *IEEE Signal Processing Letters*, 22(10):1786–1790, October 2015.
42. B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado. Subset selection in noise based on diversity measure minimization. *IEEE Trans. Signal Process.*, 51(3):760–770, March 2003.
43. A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet. Euclid in a taxicab: Sparse blind deconvolution with smoothed l1/l2 regularization. *IEEE Signal Processing Letters*, 22(5):539–543, 2015.
44. P. Rodriguez and B. Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE Trans. Image Process.*, 18(2):322–332, February 2009.
45. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
46. I. W. Selesnick, A. Parekh, and I. Bayram. Convex 1-D total variation denoising with non-convex regularization. *IEEE Signal Processing Letters*, 22(2):141–144, February 2015.
47. I.W. Selesnick and I. Bayram. Sparse signal estimation by maximally sparse convex optimization. *Signal Processing, IEEE Transactions on*, 62(5):1078–1092, March 2014.
48. S. Voronin and R. Chartrand. A new generalized thresholding algorithm for inverse problems with sparsity constraints. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, pages 1636–1640, 2013.
49. Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imag. Sci.*, 1(3):248–272, 2008.

50. Y.-B. Zhao and D. Li. Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems. *SIAM J. Optim.*, 22(3):1065–1088, 2012.