Journal of Optimization Theory and Applications (2024) 202:1286–1325
https://doi.org/10.1007/s10957-024-02479-2

Convex Predictor–Nonconvex Corrector Optimization Strategy with Application to Signal Decomposition

Laura Girometti[1] · Martin Huska[1] · Alessandro Lanza[1] · Serena Morigi[1]

Received: 25 July 2023 / Accepted: 9 June 2024 / Published online: 4 July 2024

Abstract

Many tasks in real life scenarios can be naturally formulated as nonconvex optimization problems. Unfortunately, to date, the iterative numerical methods to find even only the local minima of these nonconvex cost functions are extremely slow and strongly affected by the initialization chosen. We devise a predictor–corrector strategy that efficiently computes locally optimal solutions to these problems. An initialization-free convex minimization allows to *predict* a global good preliminary candidate, which is then *corrected* by solving a parameter-free nonconvex minimization. A simple algorithm, such as alternating direction method of multipliers works surprisingly well in producing good solutions. This strategy is applied to the challenging problem of decomposing a 1D signal into semantically distinct components mathematically identified by smooth, piecewise-constant, oscillatory structured and unstructured (noise) parts.

Keywords Nonconvex optimization · Signal decomposition · Predictor–corrector · Cross-correlation · Multi-parameter selection

Mathematics Subject Classification 65K10 · 68U99

Communicated by Anil V. Rao.

✉ Serena Morigi
serena.morigi@unibo.it

Laura Girometti
laura.girometti2@unibo.it

Martin Huska
martin.huska@unibo.it

Alessandro Lanza
alessandro.lanza2@unibo.it

[1] University of Bologna, P.zza Porta S. Donato 5, 40126 Bologna, Italy

1 Introduction

Many applications, including signal and image processing, wireless communications and machine learning, involve the solution of nonconvex and nonsmooth optimization problems. Despite their strong applicability, they are computationally challenging to solve due to the presence of multiple local stationary points that are not necessarily global optima, from the perspective of the mathematical model, and due to sensitivity to initialization, low efficiency, and critical convergence, from the algorithmic point of view.

In this work, we address the following generic nonconvex parametric optimization problem

$$x^* \in \arg\min_{x\in\mathbb{R}^N} \{\mathcal{J}(x;a,\gamma_1,\gamma_2,\ldots) := \gamma_1\mathcal{J}_1(x;a) + \gamma_2\mathcal{J}_2(x) + \gamma_3\mathcal{J}_3(x) + \cdots + \mathcal{J}_M(x)\}, \tag{1}$$

where the cost function $\mathcal{J}$ is parameterized by a set of parameters $\gamma_1, \gamma_2, \ldots$, which balance the action of M different energy terms. The function $\mathcal{J}$ is defined by the sum of a nonconvex parametric term $\mathcal{J}_1(x;a)$, which makes $\mathcal{J}(x;a,\gamma_1,\gamma_2,\ldots)$ eventually nonconvex, and other convex smooth energy terms $\mathcal{J}_2(x), \mathcal{J}_3(x), \ldots, \mathcal{J}_M(x)$.
Widely used nonconvex terms, employed for their sparsity-inducing properties, belong to the class of parameterized penalty functions $\phi(x;a)$, $a \in \mathbb{R}_+$, characterized by a lower-bounded second derivative $\inf_{x\in\mathbb{R}_{++}} \phi''(x;a) = -a$. In such a way the degree of nonconvexity can be tuned: setting $a = 0$ ensures $\mathcal{J}_1(x;a)$ to be convex, nonconvex for $a > 0$. We refer to [25] for popular examples of (sparsity-promoting) parameterized nonconvex penalty functions.

A class of popular methods developed for tackling these nonconvex optimization of eventually non-smooth functions is centered around the convex relaxation idea, which approximates the nonconvex and nonsmooth problem by a sequence of tractable convex subproblems. For the resulting convex subproblems, one then applies convex analytic tools and algorithms to compute the minimizer.

The basic approach in this direction consists in convexifying the objective function. Natural choices are the convex envelope (or biconjugate) for the objective function and convex hull for the constraint set (see, e.g., [18]), which together provide the tightest convex approximation of the original problem but might be computationally intractable. The more sophisticated and versatile version, referred to as "Convex NonConvex" (CNC) strategy [25], involves the construction and then optimization of convex functionals containing (a general class of) parameterized nonconvex sparsity-inducing separable and non-separable penalties. Not all nonconvex functionals can be convexified using CNC strategies because they lack a sufficiently convex component.

From an algorithmic perspective, "Graduated NonConvexity" (GNC) algorithms, term coined by Blake and Zisserman in [2], solve a well-constructed sequence of nonconvex problems of increasing complexity to gradually approach the target solution [31]. Along this algorithmic line of research we also mention the Majorization-Minimization (MM) algorithm, which works by iteratively optimizing a sequence of easy- to-optimize surrogate functions that bound the objective [19]. Two of the

most successful instances of MM algorithms are Expectation–Maximization (EM) [14] and the Concave–Convex Procedure (CCP) [39]. Another related line of research concerns the Difference-of-Convex (DC) programming [36], which can be shown to reduce to CCP under certain conditions.

These techniques have proved useful to a number of computer vision and learning problems [19, 31]. However, despite their widespread success, they may be very tricky to deploy in general and present a number of drawbacks in practice, some of which have motivated our work. A tractable convex surrogate may be hard to find, or suboptimal in performance and expensive in computation, and finding a good initialization poses significant practical challenges. With generic traditional, typical iterative methods at best only guarantee to converge to a critical point that might be even a saddle point and their efficiency could be seriously compromised.

The proposed numerical approach for solving (1) is inspired by the Predictor–Corrector (PC) method [4], designed to integrate ordinary differential equations, which use a suitable combination of an explicit and an implicit technique to improve the approximation accuracy while obtaining better convergence characteristics. Applying the PC strategy in the context of solving a nonconvex nonsmooth optimization problem formulated as in (1), the task of the predictor is carried out by a convex optimizer, used to predict a good approximation of the solution. The corrector, a nonconvex optimizer, is applied to improve the approximation obtained by the convex method.

A fast convex Predictor generates preliminary candidate points in the solution search space, which further allow for an automatic parameter selection for the Corrector. The initial approximation is close enough to the optimal solution to clearly facilitate and accelerate the process of solving the minimization problem by a limited number of iterations of a nonconvex Corrector.

For our purposes model (1) will assume the role of *Predictor* by setting $a = 0$, and *Corrector* for any positive a value.

The variational model (1) relies on different penalty terms in order to capture the different components of the data. While the resulting multipenalty approach has in principle a greater potential for accurate reconstructions than single-penalty models, its practical performance relies heavily on a good choice of the multiple parameters. The PC optimization strategy aims at overcoming the downsides of empirical parameter selection rules by an automatic context-aware choice of the parameter values.

Many signal/image processing problems are formulated as nonconvex variational models of the type (1) that contain a parameterized nonconvex penalty $\phi(x; a)$; examples include spike deconvolution [28], image segmentation and restoration [6, 21, 37], image decomposition [34], image inpainting [32], signal and image denoising [7, 24]. The commonly used numerical optimization strategies for solving such nonconvex models range from DC programming, when the nonconvex term can be decomposed as the difference of two convex functions, CNC strategy, when the sufficient convexity of the other convex terms is guaranteed, to GNC method, which redefines the nonconvex term in each iteration to obtain a gradually sparser solution. Due to lacks of efficient solvers, the broad applications of GNC have been limited.

In this work, the PC optimization strategy is applied to the nonconvex optimization formulation used to solve the challenging problem of decomposing a 1D signal into semantically distinct components. This problem is based on the hypothesis that an

observed signal $f \in \mathbb{R}^N$ can be represented mathematically as a mixture of three components:

$$f = c + s + o, \tag{2}$$

where $c \in \mathbb{R}^N$ is the "cartoon" approximation of f represented as a piecewise-constant (sparse-derivative) function, $s \in \mathbb{R}^N$ is a smooth low-oscillating function, and $o \in \mathbb{R}^N$ contains zero-mean high-oscillations of the signal, which can include additive white noise, as well as meaningful oscillatory components.

We propose a nonconvex variational decomposition model, which encodes qualitative features of signals into variational energies in conjunction with advanced models in sparse optimization. The three-terms variational model is analysed in terms of convexity, coercivity and existence/uniqueness of solutions in Sect. 4; it is then efficiently optimized using the Predictor–Corrector algorithmic framework (presented in Sect. 3), which incorporates an automatic multi-parameter selection based on cross-correlation principle, (described in Sect. 6). For both optimization stages the numerical optimization relies on an alternating direction method of multipliers (ADMM) approach (see Sect. 5). Numerical results (presented in Sect. 8) will show excellent performance of the proposed PC-optimizer when applied to the signal decomposition of synthetically generated signal examples, with respect to state-of-the-art variational 1D-decomposition models.

The main contributions of this paper are summarized as follows:

- a general algorithmic framework based on Predictor–Corrector strategy is introduced to tackle the minimization of nonsmooth nonconvex optimization problems in the form (1);
- a novel variational model for the ternary decomposition of 1D signals into cartoon, smooth and oscillatory components. We further devise a post-processing optimization stage able to separate oscillatory components from noise in a corrupted signal;
- an ADMM-based minimization algorithm which, thanks to a suitable variable splitting, and an adaptive step-size update, allows for a very efficient computation of the model solution with convergence guaranteed;
- a simple grid strategy for an effective, automatic selection of all the parameters of the Predictor and Corrector models for the decomposition task. In particular, this is achieved based on a novel Multi-Parameter Cross Correlation Principle.

2 Related Works on Signal Decomposition

Many application fields use time-frequency analysis and signal decomposition as a fundamental tool for quantitative and technical analysis. The first approaches to signal decomposition focused on frequency content and signal priors. In their seminal work, Huang et al. [20] introduced Empirical Mode Decomposition (EMD), an algorithmic method to detect and decompose a nonstationary nonlinear signal into principal "modes"—a signal with mostly compact supported Fourier spectrum. Among the more recent derivatives we mention IF [26], FIF and ALIF [10], as well as standard methods such as (Short Time) Fourier Transform and Wavelet-based methods such

as the synchrosqueezing in [13]. A fully variational model, named Variational Mode Decomposition (VMD), has been introduced in [15] to decompose an input signal into compact-support and band-limited modes in the spectral domain.

As an alternative, Morphological Component Analysis (MCA) is a time-domain analysis that deals with morphological diversity instead of frequency or scale information, taking into account the shape of the signal [38]. In [5] Cai et al. proposed a sparsity-enhanced signal decomposition method using MCA and wavelet transform by using the generalized minimax-concave penalty to improve the signal decomposition performance of the L1-norm regularized MCA. These methods are based on the ability to represent signal components as sparse combinations of atoms of predetermined dictionaries. In this direction, Meyers and Boyd in [30] introduce a general optimization framework for the 1D-decomposition in multiple classes. Although extremely flexible and effective, the proposed framework requires the intervention of an analyst with a preliminary knowledge of the mathematical characteristics of the signal to be extracted and of how many components are being sought.

Our variational proposal aligns with the class of MCA-based decomposition methods, elaborating semantically distinct components that can be mathematically identified as smooth-, cartoon-, oscillatory- structured, and unstructured (noise) parts. Since a unique component aggregates all the oscillatory parts of the signal, the method stands as a useful pre-processing to the time-frequency analysis.

The proposed variational model improves the nonconvex decomposition model (stage I) introduced in [8]. From the model point of view, firstly, the oscillatory component is captured by the G-norm penalty rather than its L^2-norm approximation in the negative Sobolev space $\mathcal{H}^{-1}$. Secondly, a fourth quadratic fidelity term is present in model [8] to impose soft constraints on the decomposition, whereas in this proposal the fidelity is hardly imposed. Moreover, the Predictor–Corrector proposal overcomes the problem of sensitivity to initialization, while being fully automatic, thus avoiding the critical multi-parameter selection in the model.

3 Predictor–Corrector Algorithmic Framework for Signal Decomposition

Given a nonconvex optimization problem in the form (1) the proposed general PC algorithmic framework is sketched in Algorithm 1. The role of the `Estimate_param()` procedure is to estimate the optimal values for the multi-parameters that characterize the cost function $\mathcal{J}$. This task is performed by the minimization of a context-aware function $\mathcal{E}(\gamma_1, \gamma_2, \ldots) : \mathbb{R}^{M-1} \to \mathbb{R}$ which represents properties of the specific application context. For the signal decomposition problem considered in this work, we will introduce a specific criterion $\mathcal{E}$ in Section 6, based on a cross-correlation principle. Another quite popular and effective criterion which might be plugged into the proposed general PC-optimizer framework when applied to many different signal/image processing problems is the so-called whiteness principle [22, 23].

The convex minimization problem is then uniquely solved by the `Predictor()` optimizer with fixed parameters $\widehat{\gamma}_1, \widehat{\gamma}_2, ..,$ previously evaluated by the `Estimate_param()` procedure in step [a]. The global minimizer $\widehat{y}$ plays the

role of starting point for the iterative procedure used for solving the nonconvex `Corrector()` minimization. The multi-parameter estimation (`Estimate_param()`) performed in step [c] applies a natural correction to the multi-parameters used in the `Predictor()`, and also estimates the parameter a which imposes optimal nonconvexity degree. The convergence to a local (and, possibly, global) minimizer of the original optimization problem (1) with the parameters $a, \gamma_1, \gamma_2, \ldots$ selected is therefore favored and accelerated by having a good initialization.

Algorithm 1 PC-optimizer

[a] $(\widehat{\gamma}_1, \widehat{\gamma}_2, \ldots) \leftarrow$ `Estimate_param`$(\mathcal{J}, \mathcal{E})$
[b] $(\widehat{y}) \leftarrow$ `Predictor`$(\mathcal{J}(x; a = 0, \widehat{\gamma}_1, \widehat{\gamma}_2, ..))$
[c] $(a, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2, \ldots) \leftarrow$ `Estimate_param`$(\mathcal{J}, \mathcal{E})$
[d] $(y) \leftarrow$ `Corrector`$(\mathcal{J}(\widehat{y}; a, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2, ..))$

It arises naturally the comparison with the GNC, introduced in [2], that follows an outer iterative process to compute a solution to the nonconvex problem by starting from its convex surrogate and gradually changing a (i.e., gradually increasing the amount of nonconvexity of a parametric nonconvex penalty) until the original nonconvex function is recovered. The solution obtained at each outer iteration is used as the initial guess for the subsequent iteration. The proposed PC-optimizer shares the simple idea that, when solving a nonconvex optimization problem with parameterized degree of nonconvexity, in order to reduce the problem of local minimizers, it is advantagious to estimate a good initial guess for the iterative optimization algorithm by solving a convex (PC-optimizer) or a sequence of less nonconvex (GNC) surrogate problems. However, the PC-approach improves this idea with a framework for the fully automatic estimation of all the free parameters of the optimized cost function, including the parameter a controlling the degree of nonconvexity of the model. As a result, unlike the GNC approach, our efficient proposal can be seamlessly used for practical applications where the automatic parameter selection is a must.

In the following we apply the PC optimizer to the nonconvex optimization problem formulated for the decomposition of a 1D signal into semantically homogeneous components. We point out that the applicability of the PC optimizer is not limited to the considered decomposition setup. Other important applications in variational signal and image processing can be easily treated.

We estimate the signal components cartoon c, smooth s, and oscillatory o of a given sampled signal $f \in \mathbb{R}^N$ by solving the following optimization problem

$$\{\widehat{c}, \widehat{s}, \widehat{o}\} \in \underset{c,s,o \in \mathbb{R}^N}{\arg\min} \left\{ \gamma_1 \sum_{i=1}^{N} \phi\left(|(\mathrm{D}c)_i|_2; a\right) + \frac{\gamma_2}{2} \|\mathrm{H}s\|_2^2 + \|o\|_G \right\}, \tag{3}$$

$$\text{subject to}: \; c + s + o = f, \;\; \sum_{i=1}^{N} c_i = 0, \tag{4}$$

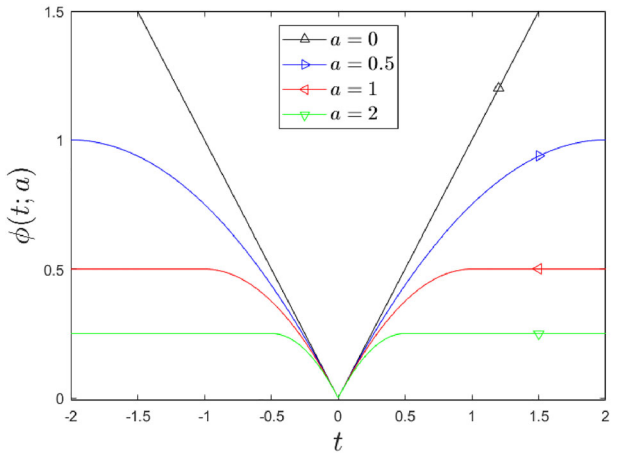

Fig. 1 The parametrized penalty function $\phi(t; a)$ used for different values of a

with penalty function ϕ defined in (5), G-norm defined in (6), and $\mathrm{D}, \mathrm{H} \in \mathbb{R}^{N\times N}$ finite difference matrices discretizing the first- and second-order derivatives. The role of the constraint on the sum of the entries of the cartoon component c is to restrict the solution space, as will be detailed in Sect. 4.

Definition 3.1 (*minimax concave penalty function*) The minimax concave (MC) penalty function $\phi : \mathbb{R} \to \mathbb{R}_+$ with parameter $a \in \mathbb{R}_+$ is defined by

$$\begin{aligned} \phi(t; a) &= \begin{cases} -\dfrac{a}{2}t^2 + |t| & \text{if } |t| \in [0, 1/a]\,, \\ \dfrac{1}{2a} & \text{if } |t| \in]1/a, +\infty[\,, \end{cases} \qquad \text{for } a \in \mathbb{R}_{++}, \\ \phi(t; 0) &= \lim_{a \searrow 0} \phi(t; a) = |t|\,, \end{aligned} \tag{5}$$

and is illustrated in Fig. 1

We recall the Meyer's G-space—introduced in [29] in the 2-dimensional domain, which is the dual of BV-space, and contains distributions f that can be written as $f = \partial_1 g_1 + \partial_2 g_2 = -\partial_1^* g_1 - \partial_2^* g_2$, for $g_1, g_2 \in L^\infty(\mathbb{R}^2)$. The G-norm $\|f\|_G$ is small for highly oscillating components characterized by zero-mean. We now restrict the G-space to the discrete 1-dimensional domain containing uniformly spaced samples of a bounded function f, denoted by the vector $o \in \mathbb{R}^N$ and endowed with the G-norm defined as

$$\|o\|_{\mathrm{G}} := \inf \left\{ \|g\|_\infty \;\middle|\; o = -\mathrm{D}^{\mathrm{T}} g, \;\; g \in \mathbb{R}^N \right\}, \tag{6}$$

where $\|g\|_\infty := \max_i |g_i|$, $\mathrm{D} \in \mathbb{R}^{N\times N}$ denotes the first-order linear differential operator, and D^{T} its transpose. Therefore, by penalizing $\|o\|_G$ we aim to capture the higher-frequency oscillatory parts of the signal characterized by zero-mean.

Both `Predictor` and `Corrector` optimizers will be applied to solve the same optimization problem (3)–(4); by setting $a = 0$ the `Predictor` solves a convex minimization problem, whereas for $a > 0$ the `Corrector` solves a nonconvex minimization problem.

The `Estimate_param()` procedure uses the ADMM-based algorithm described in Sect. 5 to approximatively solve problem (3)–(4) for a grid $[\underline{\gamma}_1, \overline{\gamma}_1] \times [\underline{\gamma}_2, \overline{\gamma}_2]$ of N_p logarithmically equispaced different values of the parameters γ_1 and γ_2. This requires to solve N_p instances of the decomposition problem and then apply an ad hoc criterium $\mathcal{E}$ to select the optimal parameter set according to the specific application the PC framework is aimed to solve. For signal decomposition problems we propose to minimize a cross-correlation function, introduced in Sect. 6, which measures the separability among the different components.

The coarse grid used in step [a] in Algorithm 1 will cover a sufficiently large feasible region, while in step [c] the refined grid will be localized near the best estimated values $\widehat{\gamma}_1, \widehat{\gamma}_2$ computed in step [a]; this results in a correction of the parameters around the values obtained for the convex Predictor solver.

4 Model Reformulation and Analysis

In this section, first we reformulate the decomposition model (3)–(4) into a suitable equivalent one with associated minimization problem of reduced dimensionality (dimension $2N$ instead of $3N$). Then, we analyze the model in terms of convexity, coercivity and existence/uniqueness of solutions.

By making use of the definition (6) of the G-norm of the oscillatory component $o = \mathrm{D}^{\mathrm{T}} g$ and replacing $s = f - c - o = f - c - \mathrm{D}^{\mathrm{T}} g$, problem (3)–(4) can be quite easily reformulated in the following equivalent form of reduced dimensionality:

$$\{\widehat{c}, \widehat{g}\} \in \underset{c,g \in \mathbb{R}^N}{\arg\min} \left\{ \widetilde{\mathcal{J}}(c, g; a, \gamma_1, \gamma_2) := \mathcal{J}(c, g; a, \gamma_1, \gamma_2) + \iota_{\mathcal{C}}(c) \right\}, \quad (7)$$

with function $\mathcal{J}$ and (linear) constraint set $\mathcal{C}$ defined by

$$\mathcal{J}(c, g; a, \gamma_1, \gamma_2) = \gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}c)_i\right|; a\right) + \frac{\gamma_2}{2} \|\mathrm{H}(c + \mathrm{D}^{\mathrm{T}} g - f)\|_2^2 + \|g\|_\infty, \quad (8)$$

$$\mathcal{C} = \left\{ c \in \mathbb{R}^N : \sum_{i=1}^{N} c_i = 0 \right\}, \quad (9)$$

and with estimated oscillatory and smooth components given by

$$\widehat{o} = \mathrm{D}^{\mathrm{T}} \widehat{g}, \quad \widehat{s} = f - \widehat{o} - \widehat{c}. \quad (10)$$

Some of the reported results (or their proof) can depend on the discretization choices for the first- and second-order differential matrices D and H in (8). However, analogous results could be obtained in a similar manner for other discretization schemes.

Here, we adopt unscaled forward finite difference discretizations for the first-order derivatives and unscaled centered finite difference discretizations for the second-order derivatives. More formally, depending on the particular boundary conditions (BC) adopted among the periodic (P), reflective (R) and anti-reflective (A) ones, the six matrices $\mathrm{D_P}, \mathrm{D_R}, \mathrm{D_A}, \mathrm{H_P}, \mathrm{H_R}, \mathrm{H_A} \in \mathbb{R}^{N\times N}$ are explicitly given by

$$\mathrm{D_P} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ 1 & 0 & \dots & 0 & -1 \end{bmatrix}, \quad \mathrm{H_P} = \begin{bmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 1 & 0 & \dots & 1 & -2 \end{bmatrix},$$

$$\mathrm{D_R} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad \mathrm{H_R} = \begin{bmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{bmatrix}, \tag{11}$$

$$\mathrm{D_A} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & -1 & 1 \end{bmatrix}, \quad \mathrm{H_A} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

In the subsequent analyses, we will consider all three types of boundary conditions, as they all are of practical interest for signal processing and, as we will see, involve differences in the theoretical analysis of the decomposition model considered. Specifically, it is well-known that the null spaces of the matrices are given by:

$$\begin{aligned} \mathrm{null}\,(\mathrm{D_{BC}}) &= \mathrm{span}\,(1_N) \quad \forall\, \mathrm{BC} \in \{\mathrm{P}, \mathrm{R}, \mathrm{A}\}\,, \\ \mathrm{null}\,(\mathrm{H_P}) = \mathrm{null}\,(\mathrm{H_R}) &= \mathrm{span}\,(1_N)\,, \quad \mathrm{null}\,(\mathrm{H_A}) = \mathrm{span}\,(1_N, l_N)\,, \end{aligned}$$

where vector l_N is defined by

$$l_N := \big(-(N-1)/2,\ -(N-1)/2+1, , \dots,\ (N-1)/2-1,\ (N-1)/2\big), \tag{12}$$

so that it holds true that

$$1_N^{\mathrm{T}}\, l_N = 0, \quad \mathrm{D_A}\, l_N = 1_N\,, \quad \forall\, N \in \mathbb{N}\,. \tag{13}$$

Moreover, it is immediate to verify that, for any vector $g \in \mathbb{R}^N$, we have

$$\sum_{i=1}^{N} o_i = \sum_{i=1}^{N} \left(\mathrm{D_{BC}^T} g\right)_i = 1_N^{\mathrm{T}} \left(\mathrm{D_{BC}^T} g\right)$$

$$= (\mathrm{D}_{\mathrm{BC}} 1_N)^{\mathrm{T}} g = 0 \quad \forall\, \mathrm{BC} \in \{\mathrm{P}, \mathrm{R}, \mathrm{A}\}, \tag{14}$$

which proves that the oscillatory component $\widehat{o}$ estimated by our discrete decomposition model—see (10)—always has the desired theoretical property of being zero-mean.

In the following Proposition 4.1 we analyze the cost function $\mathcal{J}$ in (8) and prove that it admits an infinity of global minimizers. Then, in Proposition 4.2 and Corollary 4.1 we outline the (desirable) effects of imposing, while minimizing $\mathcal{J}$, a suitable class of linear constraints, of which the one considered in our model (9) is a specific instance. The proofs are postponed to the Appendix and, to simplify notations, we introduce the total optimization variable $x := \big(c, g\big)^T \in \mathbb{R}^{2N}$.

Proposition 4.1 *For any $f \in \mathbb{R}^N$ and any $\gamma_1, \gamma_2 \in \mathbb{R}_{++}$, $a \in \mathbb{R}_+$, the function $\mathcal{J}$ in (8) is proper, continuous and bounded from below by zero. Moreover,*

(1) *$\mathcal{J}$ is constant along straight lines in its domain $\mathbb{R}^{2N}$ of direction defined by the vector*

$$d := (1_N; 0_N), \tag{15}$$

hence $\mathcal{J}$ is non-coercive in x.
(2) *for $a = 0$, $\mathcal{J}$ is convex in x.*
(3) *for $a > 0$, $\mathcal{J}$ is convex in g but nonconvex in x.*
(4) *in spite of its non-coercivity, $\mathcal{J}$ admits an infinity of global minimizers for any $a \geq 0$.*

Proposition 4.2 *Let $\mathcal{J}$ be the function in (8). Then, for any $f \in \mathbb{R}^N$ and any $\gamma_1, \gamma_2 \in \mathbb{R}_{++}$, $a \in \mathbb{R}_+$, any constrained optimization problem of the form*

$$\{\widehat{c}, \widehat{g}\} \in \underset{c,g \in \mathbb{R}^N}{\arg\min} \left\{ \widetilde{\mathcal{J}}(c, g; a, \gamma_1, \gamma_2) := \mathcal{J}(c, g; a, \gamma_1, \gamma_2) + \iota_{\mathcal{H}_v}(c, g) \right\}, \tag{16}$$

with $\mathcal{H}_v \subset \mathbb{R}^{2N}$ defined by

$$\mathcal{H}_v := \left\{ x \in \mathbb{R}^{2N} : \; v^T x = 0 \;\; \text{with} \;\; v \in \mathbb{R}^{2N} : \; v^T d \neq 0 \right\}, \tag{17}$$

is equivalent to its unconstrained counterpart in terms of minimum cost function value, but the dimensionality of the space of equivalent global minimizers is reduced by 1.

Corollary 4.1 *Let $\widetilde{\mathcal{J}}$ be the function in (16), with $\mathcal{J}$ in (8) and $\mathcal{H}_v$ in (17). Then,*

(1) *for $a = 0$, $\widetilde{\mathcal{J}}$ is coercive in x for* $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}, \mathrm{A}\}$.
(2) *for $a > 0$, $\widetilde{\mathcal{J}}$ is coercive in x for* $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}\}$.

In our framework the convex predictor is a key ingredient for the success of the overall PC optimizer. This motivated us to investigate more deeply, in the following Lemma 4.1 and Proposition 4.3 (whose proofs are postponed to the Appendix), the (constrained) predictor model in terms of existence/uniqueness of solutions.

Lemma 4.1 *Let $\mathcal{H}_v$ be the class of hyperplanar constraint sets defined in (17) with $v = (v_1, v_2)^T \in \mathbb{R}^{2N}$, $v_1, v_2 \in \mathbb{R}^N$, and let $\mathrm{Q_{BC}} \in \mathbb{R}^{N\times 2N}$ be the matrix given by*

$$\mathrm{Q_{BC}} := [\,\mathrm{H_{BC}},\ \mathrm{H_{BC}D_{BC}^T}\,] = \mathrm{H_{BC}}\,[\,\mathrm{I}_N\,,\ \mathrm{D_{BC}^T}\,],\quad \mathrm{BC} \in \{\mathrm{P},\mathrm{R},\mathrm{A}\}, \tag{18}$$

with $\mathrm{D_{BC}}, \mathrm{H_{BC}} \in \mathbb{R}^{N\times N}$ the first- and second-order finite difference matrices defined in (11), depending on the boundary conditions BC*. Then, it holds true that*

$$\mathcal{N}_{\mathrm{BC}} := \mathrm{null}(\mathrm{Q_{BC}}) \cap \mathcal{H}_v = \mathrm{range}\,(\mathrm{M_{BC}})\ , \tag{19}$$

with matrix $\mathrm{M_{BC}}$ defined by

$$\mathrm{M_{BC}} = \begin{cases} \begin{bmatrix} -\mathrm{D_{BC}^T} + \frac{1}{v^T d} 1_N \left(v_1^T \mathrm{D_{BC}^T} - v_2^T\right) \\ \mathrm{I}_N \end{bmatrix}, & \text{for BC} \in \{\mathrm{P},\mathrm{R}\}, \\ \begin{bmatrix} -\mathrm{D_{BC}^T} + \frac{1}{v^T d} 1_N \left(v_1^T \mathrm{D_{BC}^T} - v_2^T\right) & \left(l_N - \left(v_1^T l_N\right) 1_N\right) \\ \mathrm{I}_N & 0_N \end{bmatrix}, & \text{for BC} = \mathrm{A}, \end{cases} \tag{20}$$

and with

$$r_{\mathrm{BC}} := \mathrm{rank}(\mathrm{M_{BC}}) = \begin{cases} N, & \text{for BC} \in \{\mathrm{P},\mathrm{R}\}, \\ N+1, & \text{for BC} = \mathrm{A}. \end{cases} \tag{21}$$

Proposition 4.3 *Let $\mathcal{J}$ be the function in (8) and $\mathcal{H}_v$ the class of hyperplanar feasible sets defined in (17). Then, in case $a = 0$, for any $f \in \mathbb{R}^N$ and any $\gamma_1, \gamma_2 \in \mathbb{R}_{++}$, the constrained optimization problem in (16) is convex and admits a m-dimensional convex compact set of solutions $\widehat{x} = (\widehat{c}, \widehat{g})^T \in \mathcal{M} \subset \mathcal{H}_v \subset \mathbb{R}^{2N}$. In particular, if the solution set $\mathcal{M}$ is not a singleton ($m > 0$), then it must belong to a m-dimensional affine subspace parallel to the linear subspace $\mathcal{N}_{\mathrm{BC}}$ in (19), hence the dimensionality of $\mathcal{M}$ satisfies $m \le r_{\mathrm{BC}}$ with r_{BC} given in (21).*

Remark 4.1 Under the assumption that $\mathcal{M}$ is either a singleton ($m = 0$) or full-dimensional ($m = 2N - 1$), then the constrained optimization problem in (16) with $a = 0$ admits a unique solution. This comes from adapting the proof in [1] to our original model (3)–(4).

Remark 4.2 We remark that the constraint set $\sum_i c_i = 0$ used in our decomposition model (7)–(10) can be seen as a particular hyperplane of the class $\mathcal{H}_v$ defined in (17), obtained for $v = d$. Hence, our model exhibits all properties stated in Proposition 4.2, Corollary 4.1, Proposition 4.3 and Remark 4.1. However, our special choice in (9) will allow to enforce the constraint in a very efficient (if not immediate) manner within one of the ADMM minimization subproblems.

5 ADMM-Based Numerical Solution

In this section, we illustrate an ADMM-based numerical solver for the proposed signal decomposition model (7)–(10), which is shared—with a single difference which

will be adequately highlighted—by both the (convex) Predictor and the (nonconvex) Corrector steps in the PC optimizer. We chose ADMM as the optimization algorithm for different reasons. First, ADMM is known to be able to provide, at least in the convex optimization case, "medium-quality" solutions in fast time. This is of crucial importance for our (convex) Predictor phase, where we aim at computing efficiently a sufficiently good-quality predicted initial guess for the (nonconvex) Corrector step. Then, thanks to suitable preliminary variable splitting choices—which will be made explicit in Sect. 5.1—the two-blocks ADMM schemes that we apply to the solution of the Predictor and Corrector models have both guaranteed convergence. In particular, in the convex Predictor case, we can also apply the very effective iteration-adaptive step-size rule quite recently proposed in [27], which makes the achievement of medium-quality solutions even faster, independently of the two model parameters γ_1, γ_2. Moreover, in the nonconvex Corrector case, we are able to demonstrate convergence of our ADMM scheme based on the recent results reported in [35].

In Sect. 5.1 we describe in detail the proposed two-blocks ADMM iterative solver, then in Sect. 5.2 we discuss about its convergence both in the convex (with adaptive step-size rule) and in the nonconvex case.

5.1 ADMM-Based Iterative Solver

To solve the minimization problem (7)–(9) by means of a standard two-blocks ADMM approach [3], first we introduce the auxiliary variable $z \in \mathbb{R}^{2N}$ defined by

$$z = x, \quad \text{with} \quad z := \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad x := \begin{pmatrix} c \\ g \end{pmatrix}, \quad z_1, z_2, c, g \in \mathbb{R}^N. \tag{22}$$

Problem (7)–(9) can thus be equivalently reformulated as

$$\{\widehat{x}, \widehat{z}\} \in \underset{x,z \in \mathbb{R}^{2N}}{\arg\min} \{ F(x) + G(z) \}, \quad \text{subject to} \quad x - z = 0_{2N}, \tag{23}$$

where functions $F, G : \mathbb{R}^N \to \mathbb{R}_+$ are defined by

$$F(x) = \frac{\gamma_2}{2} \|\mathrm{Q}x - \mathrm{H}f\|_2^2, \quad G(z) = \gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}z_1)_i\right|; a\right) + \iota_{\mathcal{C}}(z_1) + \|z_2\|_\infty, \tag{24}$$

with matrix Q given in (18).

The augmented Lagrangian function associated to problem (23) reads as

$$\mathcal{L}_\beta(x, z, \lambda) = F(x) + G(z) + \langle \lambda, x - z \rangle + \frac{\beta}{2} \|x - z\|_2^2, \tag{25}$$

where $\lambda \in \mathbb{R}^{2N}$ is the vector of Lagrange multipliers associated to the system of linear constraints $x - z = 0_{2N}$ and $\beta \in \mathbb{R}_{++}$ is the ADMM penalty parameter.

Given the previously computed (or initialized for $k = 0$) vectors $x^{(k)}, z^{(k)}, \lambda^{(k)}$, the k-th iteration of the proposed ADMM-based iterative scheme reads as follows:

$$z^{(k+1)} \in \underset{z\in\mathbb{R}^{2N}}{\arg\min}\ \mathcal{L}_\beta\big(x^{(k)}, z, \lambda^{(k)}\big), \tag{26}$$

$$x^{(k+1)} = \underset{x\in\mathbb{R}^{2N}}{\arg\min}\ \mathcal{L}_\beta\big(x, z^{(k+1)}, \lambda^{(k)}\big), \tag{27}$$

$$\lambda^{(k+1)} = \lambda^{(k)} + \beta\big(x^{(k+1)} - z^{(k+1)}\big). \tag{28}$$

The subproblem for $x^{(k+1)}$ in (27) reads as follows

$$x^{(k+1)} = \underset{x\in\mathbb{R}^{2N}}{\arg\min}\left\{\frac{\gamma_2}{2}\|\mathrm{Q}x - \mathrm{H}f\|_2^2 + \langle\lambda^{(k)}, x - z^{(k+1)}\rangle + \frac{\beta}{2}\|x - z^{(k+1)}\|_2^2\right\} \tag{29}$$

The first-order optimality conditions for (29) read

$$\left(\overbrace{\mathrm{Q}^T\mathrm{Q} + \frac{\beta}{\gamma_2}\mathrm{I}_{2N}}^{\mathrm{C}}\right)x = \mathrm{Q}^T\mathrm{H}f + \frac{\beta}{\gamma_2}\left(z^{(k+1)} - \frac{1}{\beta}\lambda^{(k)}\right). \tag{30}$$

The coefficient matrix $\mathrm{C} \in \mathbb{R}^{2N\times 2N}$ of linear system (30) is clearly symmetric and positive definite, hence the updated vector $x^{(k+1)}$ in (27) is given by the unique solution of (30). Moreover, C is very sparse, having only 19 diagonals containing non-zero elements, independently of its order $2N$—that is, of the size N of the decomposed signals, with $N \geq 6$—and of the boundary conditions $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}, \mathrm{A}\}$ assumed. Hence it is computationally convenient to compute the sparse Cholesky factorization once and for all at the beginning and then solve two triangular systems at each ADMM iterations.

The subproblem for $z^{(k+1)}$ takes the following form:

$$z^{(k+1)} \in \underset{z\in\mathbb{R}^{2N}}{\arg\min}\ \left\{G(z) + \frac{\beta}{2}\left\|z - q^{(k+1)}\right\|_2^2\right\}, \quad q^{(k+1)} = x^{(k)} + \frac{1}{\beta}\lambda^{(k)}. \tag{31}$$

Since the cost function in (31) is separable in z_1, z_2—see the definition of variable z in (22) and of function G in (24)—by introducing $\lambda_1^{(k)}, \lambda_2^{(k)} \in \mathbb{R}^N$ such that $\lambda^{(k)} = (\lambda_1^{(k)}, \lambda_2^{(k)})^T$, problem (31) is equivalent to solve the following two subproblems:

$$z_1^{(k+1)} \in \underset{z_1\in\mathbb{R}^N}{\arg\min}\ \left\{\frac{1}{2}\left\|z_1 - q_1^{(k+1)}\right\|_2^2 + \frac{\gamma_1}{\beta}\sum_{i=1}^{N}\phi\left(\left|(\mathrm{D}z_1)_i\right|; a\right) + \iota_{\mathcal{C}}(z_1)\right\}, \tag{32}$$

$$z_2^{(k+1)} = \underset{z_2\in\mathbb{R}^N}{\arg\min}\ \left\{\frac{1}{2}\left\|z_2 - q_2^{(k+1)}\right\|_2^2 + \frac{1}{\beta}\|z_2\|_\infty\right\}, \tag{33}$$

where $q_1^{(k+1)} = c^{(k)} + \frac{1}{\beta}\lambda_1^{(k)}$ and $q_2^{(k+1)} = g^{(k)} + \frac{1}{\beta}\lambda_2^{(k)}$.

Subproblem (33) is solved by first noting that $z_2^{(k+1)}$ in (33) is given by the (unique) value of the proximity operator of proper lsc convex function $\frac{1}{\beta}\|\cdot\|_\infty$ at $q_2^{(k+1)}$ and then applying the Moreau decomposition to this operator, which leads to

$$z_2^{(k+1)} = \operatorname{prox}_{\frac{1}{\beta}\|\cdot\|_\infty}\left(q_2^{(k+1)}\right) = q_2^{(k+1)} - \operatorname{proj}_{\|\cdot\|_1 \leq \frac{1}{\beta}}\left(q_2^{(k+1)}\right), \tag{34}$$

where $\operatorname{proj}_S(v)$ denotes the Euclidean projection of vector v onto the set S. Hence, based on [12], the updating problem in (34) admits a closed form efficient solution.

Finally, subproblem (32) is efficiently solved based on the results reported in the following Proposition 5.1. We remark that this is the unique subproblem where a difference between the convex ($a = 0$) and the nonconvex ($a > 0$) cases arises.

Proposition 5.1 *Let ϕ be the penalty function defined in (5) and $\mathcal{C}$ the constraint set in (9). Then, for any $\gamma_1, \beta \in \mathbb{R}_{++}$, any $a \in \mathbb{R}_+$ and any vector $q_1^{(k+1)} \in \mathbb{R}^N$, the solution(s) of the constrained minimization problem in (32) is given by*

$$\begin{aligned} z_1^{(k+1)} \in \underset{z_1 \in \mathbb{R}^N}{\arg\min} & \left\{ \frac{\gamma_1}{\beta} \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}z_1)_i\right|; a\right) + \frac{1}{2}\left\| z_1 - \widetilde{q}_1^{(k+1)} \right\|_2^2 \right\}, \\ \text{with} \quad & \widetilde{q}_1^{(k+1)} = q_1^{(k+1)} - \frac{1}{N}\sum_{i=1}^{N}\left(q_1^{(k+1)}\right)_i . \end{aligned} \tag{35}$$

Moreover, for any $q_1^{(k+1)} \in \mathbb{R}^N$, if the following condition is satisfied,

$$0 \leq a \leq \frac{\beta}{4\gamma_1}, \tag{36}$$

then the cost function in (35) is strongly convex, hence (35) admits a unique solution which for $a = 0$ can be calculated exactly in finite time [11], while for $a > 0$ can be obtained as the limit point of the following, convergent forward-backward splitting (FBS) iterative algorithm initialized by $\theta^{(0)}$:

$$\theta^{(j+1/2)} = a\,\mathrm{D}^T\left(\mathrm{D}\theta^{(j)} - \operatorname{soft}_{1/a}\left(\mathrm{D}\theta^{(j)}\right)\right), \tag{37}$$

$$\theta^{(j+1)} = \underset{\theta \in \mathbb{R}^N}{\arg\min}\left\{ \frac{\gamma_1}{\beta}\|\mathrm{D}\theta\|_1 + \frac{1}{2}\left\| \theta - \left(\tilde{q}_1^{(k+1)} + \frac{\gamma_1}{\beta}\theta^{(j+1/2)}\right)\right\|_2^2 \right\}. \tag{38}$$

The proof of (35) is immediate, whereas a detailed demonstration of (36)–(38) is given in [33]. We recall that, in the forward step (37), the soft threshold function $\operatorname{soft}_T : \mathbb{R} \to \mathbb{R}$ with threshold parameter $T \geq 0$ is defined as

$$\operatorname{soft}_T(t) = \begin{cases} 0 & |t| \leq T, \\ (|t| - T)\operatorname{sign}(t) & |t| \geq T. \end{cases} \tag{39}$$

We observe that, when $a = 0$ (which is our Predictor case), subproblem (35) reduces to a standard 1-dimensional TV-ℓ_2 denoising problem which, as stated in Proposition 5.1, can be solved exactly in finite-time by means of the very efficient approach proposed in [11]. When $a > 0$ (which is our Corrector case), the iterative FBS algorithm in (37)–(38) is initialized by $\theta^{(0)} = z_1^{(k)}$, the solution of the same subproblem obtained at the previous ADMM iteration. This represents a good initial guess and, thus, accelerates convergence. Iterations of the FBS algorithm are stopped as soon as the iterates relative change defined by $\|\theta^{(j+1)} - \theta^{(j)}\|_2 / \|\theta^{(j)}\|_2$ drops below a threshold. In particular, the threshold is decreased along the (outer) ADMM iterations until it reaches the (very low) value 10^{-16}. This guarantees that, after a reasonable number of ADMM iterations, the z_1 subproblem is solved (almost) exactly also in the case $a > 0$.

5.2 Convergence Remarks

Performance of the two-blocks ADMM algorithm crucially depends on the choice of the dual step-size β. More precisely, in the convex setting under standard assumptions on the two cost functions $F(x)$, $G(z)$ and on the linear constraint $\mathrm{A}x + \mathrm{B}z = b$, the step-size β affects only the speed of convergence in the nonconvex setting, β can also affect convergence of the iterative scheme. Proving convergence of the two-blocks ADMM in the nonconvex case under general assumptions is still an open issue.

For what regards the proposed ADMM scheme applied to the convex Predictor model, we follow the results in [27]. More precisely, in (28) we use an iteration-adaptive step-size β_k which, according to the proposal in [27], is updated at each ADMM iteration after (28) as follows:

$$\beta_{k+1} = (1 - \omega_k)\beta_k + \omega_k \mathrm{proj}_{[\beta_m, \beta_M]} \left(\frac{\|\lambda^{(k+1)}\|_2}{\|z^{(k+1)}\|_2} \right), \quad \omega_k \in (0, 1), \tag{40}$$

with $0 < \beta_m < \beta_M < \infty$, and $\omega_k \in (0, 1]$ a summable "conservation sequence" with $\omega_0 = 1$. As a consequence, the sequence $\{\lambda_k\}$ generated by the ADMM scheme using the dual step-size updating rule in (40) weakly converges to a solution of the dual problem, then the ADMM convergence is guaranteed, as outlined in [27]. These conditions are guaranteed by explicitly adding constraints to the stepsize. As already observed in [27] applying the proposed step-size rule improves the convergence rate. This leads to a more efficient `Estimate_param()` procedure which relies on a large grid for the multi-parameter search of the convex Predictor.

Next, the convergence of two-blocks ADMM scheme for $a > 0$ (nonconvex case) is guaranteed by [Theorem 5.6, Theorem 5.8 [35]], that are summarized in the following theorem:

Theorem 5.1 *Consider the following linearly constrained, two-blocks separable minimization problem*

$$\min_{(x,z)\in\mathbb{R}^m\times\mathbb{R}^n} \{ F(x) + G(z) \}, \qquad \textit{subject to} \quad Ax + Bz = b, \tag{41}$$

with $F : \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup +\infty$, $A \in \mathbb{R}^{p\times m}$, $B \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$, *and assume that:*

1 *Functions* F *and* G *are both proper and lower semi-continuous (lsc).*
2 *Matrix* A *is onto (has full row rank).*
3 *Function* $\varphi_F : \mathbb{R}^p \to \overline{\mathbb{R}}$ *defined by* $\varphi_F(s) := \inf_{x\in\mathbb{R}^m} \{F(x) \mid Ax = s\} \in C^{1,1}(\mathbb{R}^p)$ *is* L_{φ_F}*-smooth, hence* σ_{φ_F}*-hypoconvex with* $|\sigma_{\varphi_F}| \leq L_{\varphi_F}$.
4 *Function* $\varphi_G : \mathbb{R}^p \to \overline{\mathbb{R}}$ *defined by* $\varphi_G(s) := \inf_{z\in\mathbb{R}^n} \{G(z) \mid Bz = s\}$ *is lower semi-continuous.*
5 *Problem (41) has a solution:* $\arg\min \Phi \neq \emptyset$, *where*

$$\Phi(x, z) := F(x) + G(z) + \iota_{Ax+Bz=b}(x, z).$$

Then, by considering a sequence $\{x^{(k)}, z^{(k)}, \lambda^{(k)}\}_{k\in\mathbb{N}}$ *generated by a two-block ADMM scheme starting from any initial guess* $\{x^{(0)}, z^{(0)}, \lambda^{(0)}\}$ *with* $\beta > L_{\varphi_F}$, *the following hold:*

(i) (subsequential convergence) all the cluster points $\{x^*, z^*, \lambda^*\}$ *of sequence* $\{x^{(k)}, z^{(k)}, \lambda^{(k)}\}_{k\in\mathbb{N}}$ *satisfy the KKT conditions of problem (41) and attain the same finite cost function value* $F(x^*) + G(z^*)$, *this being the limit of the scalar sequence* $\{\mathcal{L}_\beta(x^{(k)}, z^{(k)}, \lambda^{(k)})\}_{k\in\mathbb{N}}$;
(ii) (boundedness) the sequence $\{Ax^{(k)}, Bz^{(k)}, \lambda^{(k)}\}_{k\in\mathbb{N}}$ *is bounded provided that the cost function* Φ *is level bounded. If, additionally,* $F \in C^{1,1}(\mathbb{R}^m)$, *then the sequence* $\{x^{(k)}, z^{(k)}, \lambda^{(k)}\}_{k\in\mathbb{N}}$ *is bounded.*
(iii) (global convercence) if Φ *is level bounded and* F *and* G *are semialgebraic, sequence* $\{Ax^{(k)}, Bz^{(k)}, \lambda^{(k)}\}_{k\in\mathbb{N}}$ *is convergent.*

Corollary 5.1 *Consider the minimization problem in (23) with* F, G *defined in (24). Then, the ADMM scheme (27)–(28) has guaranteed global convergence for* $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}\}$ *and (at least) subsequential convergence for* $\mathrm{BC} = \mathrm{A}$.

Proof It's immediate to verify that assumptions $1 - 5$ of Theorem 5.1 are satisfied. In fact, F, G defined in (24) are proper and continuous hence lsc; minimization problem (23) can be reformulated as problem (41) with $A = I_{2N}$, $B = -I_{2N}$ and $b = 0_{2N}$ and thus $A = I_{2N}$ is surjective, $\varphi_F = F$ and $\varphi_G = G$. Hence, since F is a quadratic convex function and G is lsc, φ_F is $C^{1,1}(\mathbb{R}^p)$ and $L_F - smooth$ and φ_G is lsc. Finally, problem (23) is equivalent to problem (7)–(9) that admits a solution, as outlined in Proposition 4.2. Consequently, the subsequential convergence of ADMM scheme (27)–(28) is guaranteed for all boundary conditions $\mathrm{BC} = \{\mathrm{P}, \mathrm{R}, \mathrm{A}\}$. To prove global convergence, we need to verify that Φ is level bounded and F, G are semialgebraic. The level-boundedness of function Φ holds true only for $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}\}$, as outlined in statement 2) of Corollary 4.1, where function $\widetilde{\mathcal{J}}$ defined in (16) is proven to be coercive and, hence, level bounded yielding level-boundedness of function Φ. The semialgebraicity of F and G is satisfied. In fact, F is a quadratic function while G is the sum of three semialgebraic functions: a polyhedral norm, the indicator function of a system of linear constraints and a piecewise quadratic function. Hence, the global convergence is guaranteed for $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}\}$.

In conclusion, the overall PC-Optimizer in Algorithm 1 performing a two-block ADMM optimization for convex predictor and a two-block ADMM for nonconvex corrector has guaranteed global convergence for BC $\in$ {P, R} and (at least) subsequential convergence for BC = A.

6 Multi-parameter Selection

Core of the estimation procedure `Estimate_param()` is the multi-parameter selection for the optimal balance of the energy terms in the cost function $\mathcal{J}$ which will be formulated as a minimization problem based on the cross-correlation criterium between estimated components. This allows to select the parameters γ by imposing the minimal cross-correlation between the components thus yielding the maximum dissimilarity between them.
To this aim, let us recall the definition of sample normalized cross-correlation $\rho(x, y)$ between two non-zero signals in vector form $x, y \in \mathbb{R}^N$ defined on $\Omega := \{0, \ldots, N-1\}$. The vector-valued function $\rho : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ is defined as

$$\rho(x, y) = \{\rho_l(x, y)\}_{l\in\Omega}, \quad \rho_l(x, y) = \frac{1}{\|x\|_2\|y\|_2}\sum_{i\in\Omega} x_i\, y_{i+l}, \; l \in \Omega, \tag{42}$$

where the y_{i+l} value in (42) depends of the boundary conditions considered, and $\rho_l(x, y) \in [-1, 1]$ for all lags l. We introduce the following non-negative scale-independent scalar measure of correlation $\mathtt{C} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}_+$ between the signals x and y:

$$\mathtt{C}(x, y) := \|\rho(x, y)\|_\infty \; . \tag{43}$$

We propose an extension of the Cross Correlation Principle proposed in [16] named Multi-Parameter Cross Correlation Principle (MPCCP). The extension is based on the idea that the correlation between signals involves the correlation between the derivatives (considering at least the first and the second) of the signals themselves. Specifically, the multi-parameter cross-correlation scalar measure $C : \mathbb{R}^2_{++} \to \mathbb{R}_+$ reads as

$$\begin{aligned} C(\gamma_1, \gamma_2) = {} & \mathtt{C}\left(\widehat{c}_{(\gamma_1,\gamma_2)}, \widehat{o}_{(\gamma_1,\gamma_2)}\right) + \mathtt{C}\left(\widehat{s}_{(\gamma_1,\gamma_2)}, \widehat{o}_{(\gamma_1,\gamma_2)}\right) + \mathtt{C}\left(\widehat{c}_{(\gamma_1,\gamma_2)}, \widehat{s}_{(\gamma_1,\gamma_2)}\right) \\ & + \mathtt{C}\left(\mathrm{D}\widehat{c}_{(\gamma_1,\gamma_2)}, \mathrm{D}\widehat{o}_{(\gamma_1,\gamma_2)}\right) + \mathtt{C}\left(\mathrm{D}\widehat{s}_{(\gamma_1,\gamma_2)}, \mathrm{D}\widehat{o}_{(\gamma_1,\gamma_2)}\right) + \mathtt{C}\left(\mathrm{D}\widehat{c}_{(\gamma_1,\gamma_2)}, \mathrm{D}\widehat{s}_{(\gamma_1,\gamma_2)}\right) \\ & + \mathtt{C}\left(\mathrm{H}\widehat{c}_{(\gamma_1,\gamma_2)}, \mathrm{H}\widehat{o}_{(\gamma_1,\gamma_2)}\right) + \mathtt{C}\left(\mathrm{H}\widehat{s}_{(\gamma_1,\gamma_2)}, \mathrm{H}\widehat{o}_{(\gamma_1,\gamma_2)}\right) \\ & + \mathtt{C}\left(\mathrm{H}\widehat{c}_{(\gamma_1,\gamma_2)}, \mathrm{H}\widehat{s}_{(\gamma_1,\gamma_2)}\right), \end{aligned} \tag{44}$$

with $\widehat{c}_{(\gamma_1,\gamma_2)}$, $\widehat{s}_{(\gamma_1,\gamma_2)}$ and $\widehat{o}_{(\gamma_1,\gamma_2)}$ the (γ_1, γ_2)-dependent solution components of the minimization problem (3)–(4), with the cross-correlation scalar measure $\mathtt{C}(\cdot, \cdot)$ defined in (43).

Then the MPCCP applied in `Estimate_param()` in Algorithm 1 is formulated as follows:

$$\text{Select } (\gamma_1, \gamma_2) = (\widehat{\gamma}_1, \widehat{\gamma}_2), \quad \text{such that } \{\widehat{\gamma}_1, \widehat{\gamma}_2\} \in \underset{\gamma_1, \gamma_2 \in \mathbb{R}_{++}}{\arg\min} \; C(\gamma_1, \gamma_2). \tag{45}$$

The cost function $C(\gamma_1, \gamma_2)$ to be minimized in (44)–(45) can be evaluated for any given pair (γ_1, γ_2) of a grid $[\underline{\gamma}_1, \overline{\gamma}_1] \times [\underline{\gamma}_2, \overline{\gamma}_2]$. However, as it does not have an explicit form, the minimizer is computed by a simple grid-search strategy which, however, is sufficient to make the `Estimate_param()` procedure fully automatic.

In order to make the limits of the parameter grid independent on the range of the signal f a typical pre-processing—common practice when performing some type of analysis on a signal f—is to rescale or normalize the signal magnitude. However, it is not always obvious that this leads to an equivalent variational model. Under appropriate simple parameter adjustments the following results show that this holds for the proposed variational decomposition model.

Lemma 6.1 *Let $\mathcal{J}$ be the cost function defined in (8) and let $\tilde{f} := \alpha f$, with $\alpha \in \mathbb{R}_{++}$. Then, for any $c, g, f \in \mathbb{R}^N$, any $\gamma_1, \gamma_2 \in \mathbb{R}_{++}$ and any $a \in \mathbb{R}_+$, it holds true that*

$$\mathcal{J}(c, g; f, \gamma_1, \gamma_2, a) = \frac{1}{\alpha} \mathcal{J}(\tilde{c}, \tilde{g}; \tilde{f}, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{a}), \tag{46}$$

with $\tilde{c} := \alpha c$, $\tilde{g} := \alpha g$, $\tilde{\gamma}_1 = \gamma_1$, $\tilde{\gamma}_2 = \frac{\gamma_2}{\alpha}$, $\tilde{a} = \frac{a}{\alpha}$.

Proof First, it can be easily shown that the ϕ penalty function in (5) satisfies

$$\phi(\kappa t; a) = \kappa \, \phi(t; \kappa a) \quad \forall a \in \mathbb{R}_+, \; \forall \kappa \in \mathbb{R}_{++}. \tag{47}$$

In fact, for $a = 0$, the proof is immediate. For $a \in \mathbb{R}_{++}$, we have

$$\begin{aligned} \phi(\kappa t; a) &= \begin{cases} -\dfrac{a\kappa^2}{2} t^2 + \kappa |t| & \text{if } \kappa |t| \in [0, 1/a], \\ \dfrac{1}{2a} & \text{if } \kappa |t| \in]1/a, +\infty[, \end{cases} \\ &= \kappa \begin{cases} -\dfrac{\kappa a}{2} t^2 + |t| & \text{if } |t| \in [0, 1/(\kappa a)], \\ \dfrac{1}{2(\kappa a)} & \text{if } |t| \in]1/(\kappa a), +\infty[, \end{cases} = \kappa \, \phi(t; \kappa a). \end{aligned}$$

Then, we have that

$$\begin{aligned} \mathcal{J}(c, g; f, \gamma_1, \gamma_2, a) &= \gamma_1 \sum_{i=1}^{N} \phi\left(\frac{1}{\alpha} \left| (\mathrm{D}(\alpha c))_i \right| ; a \right) \\ &\quad + \frac{\gamma_2}{2} \left\| \frac{1}{\alpha} \mathrm{H}(\alpha c + \mathrm{D}^{\mathrm{T}}(\alpha g) - \alpha f) \right\|_2^2 + \left\| \frac{(\alpha g)}{\alpha} \right\|_\infty \\ &= \gamma_1 \sum_{i=1}^{N} \phi\left(\frac{1}{\alpha} \left| (\mathrm{D}\tilde{c})_i \right| ; a \right) + \frac{\gamma_2}{2\alpha^2} \left\| \mathrm{H}\left(\tilde{c} + \mathrm{D}^{\mathrm{T}} \tilde{g} - \tilde{f} \right) \right\|_2^2 + \frac{1}{\alpha} \|\tilde{g}\|_\infty \end{aligned}$$

$$
\begin{aligned}
&= \frac{\gamma_1}{\alpha} \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}\tilde{c})_i\right| ; \frac{a}{\alpha}\right) + \frac{(\gamma_2/\alpha)}{2\alpha} \left\| \mathrm{H}\left(\tilde{c} + \mathrm{D}^{\mathrm{T}}\tilde{g} - \tilde{f}\right)\right\|_2^2 + \frac{1}{\alpha} \|\tilde{g}\|_\infty \\
&= \frac{1}{\alpha}\left[\gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}\tilde{c})_i\right| ; \tilde{a}\right) + \frac{\tilde{\gamma}_2}{2} \left\| \mathrm{H}\left(\tilde{c} + \mathrm{D}^{\mathrm{T}}\tilde{g} - \tilde{f}\right)\right\|_2^2 + \|\tilde{g}\|_\infty \right] \\
&= \frac{1}{\alpha} \mathcal{J}\left(\tilde{c}, \tilde{g}; \tilde{f}, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{a}\right).
\end{aligned}
$$

□

7 Denoising of the Oscillatory Term

In this section, we present an original post-processing extension of the three components signal decomposition model (3)–(4) in order to deal with the case of noisy signals. Indeed, the noise can be viewed as a very highly oscillatory function (this means that noise is in G-space). Therefore the PC-algorithm incorporates the noise in the oscillatory component (o) and does not discriminate between the structured high-frequency components (t) and the additive noise (n): $o = t + n$.

In order to tackle signals perturbed by realizations of white noise, we introduce the sample normalized auto-correlation of a signal, the vector-valued function $\varphi : \mathbb{R}^N \to \mathbb{R}^N$, with scalar components $\varphi_l(x) = \rho_l(x, x) \in [-1, 1]$, with the normalized cross-correlation $\rho_l(x, x)$ defined in (42). Motivated by the properties of the asymptotic distribution of φ_l of a white random process at any nonzero lag l [17], the white noise component n is constrained to belong to the parametric set $\mathcal{W}_\alpha$, referred to as the *normalized whiteness set* with confidence $\alpha \in \mathbb{R}_{++}$ called the *whiteness parameter*, and defined as

$$
\begin{aligned}
\mathcal{W}_\alpha &:= \left\{ n \in \mathbb{R}^N : \ |\varphi_l(n)| \le w_\alpha, \quad \forall l = 1, \ldots, N-1 \right\} \\
&= \left\{ n \in \mathbb{R}^N : \ -w_\alpha\, n^T n \le n^T\, T_l\, n \le w_\alpha\, n^T n, \quad \forall l = 1, \ldots, N-1 \right\}, \quad (48)
\end{aligned}
$$

where, for each nonzero lag l, $\varphi_l(n)$ can be rewritten as the scalar product between n and $T_l n$, where $T_l \in \mathbb{R}^{N\times N}$ are permutation matrices associated with the following cyclic permutations

$$
\sigma_l(n_i) = n_{i+l} \ \text{ modulo } \ N, \quad \text{for } \ i \in \Omega. \tag{49}
$$

Therefore, T_l maps an input vector $n \in \mathbb{R}^N$ to

$$
(T_l\, n)_i = n_{i+l} \ i = 0, \ldots, N-1. \tag{50}
$$

A natural choice for w_α, which measures the accuracy of belonging to the whiteness set $\mathcal{W}_\alpha$, follows by the results of the asymptotic distribution of the sample normalised auto- correlation $\varphi(n)$ (see [17]), and reads as

$$
w_\alpha = \frac{\alpha}{\sqrt{N}}. \tag{51}
$$

In order to further decompose o into separate components t and n and effectively obtain denoised components, we propose the following variational model based on the whiteness-based approach:

$$\{\widehat{t},\widehat{n}\} \in \underset{t,\,n\in\mathbb{R}^N}{\arg\min} \left\{\frac{1}{2}\|\mathrm{D}t\|_2^2 + \iota_{\mathcal{W}_\alpha}(n)\right\}, \quad \text{subject to}: \; t+n=o, \tag{52}$$

where we denoted by $\iota_{\mathcal{W}_\alpha}$ the indicator function of the whiteness set $\mathcal{W}_\alpha$ defined in (48), which imposes that the component n resembles as much as possible the underlying additive noise in terms of whiteness.

An approximate solution of (52) is obtained by means of a two-block ADMM approach. We introduce one auxiliary variable $r = o - t$, the associated vector of Lagrange multipliers $\lambda_r \in \mathbb{R}^N$ and the ADMM penalty parameter $\beta_r \in \mathbb{R}_{++}$. Starting from $\left(t^{(0)}, r^{(0)}, \lambda_r^{(0)}\right)$, the resulting iterative scheme reads as follows:

$$t^{(k+1)} \in \underset{t\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathrm{D}t\|_2^2 + \frac{\beta_r}{2}\left\|t - \left(o - r^{(k)} + \frac{\lambda_r^{(k)}}{\beta_r}\right)\right\|_2^2, \tag{53}$$

$$r^{(k+1)} \in \underset{r\in\mathbb{R}^N}{\arg\min} \frac{\beta_r}{2}\left\|r - \left(o - t^{(k)} + \frac{\lambda_r^{(k)}}{\beta_r}\right)\right\|_2^2, \tag{54}$$

$$\text{s.t.} \; -w_\alpha r^T r \le r^T T_l r \le w_\alpha r^T r, \quad l = 1,\ldots,N-1, \tag{55}$$

$$\lambda_r^{(k+1)} = \lambda_r^{(k)} + \beta_r\left(r^{(k+1)} - \left(o - t^{(k+1)}\right)\right). \tag{56}$$

The solution of t-subproblem is calculated solving the symmetric positive definite linear system (57) resulting from imposing the first-order optimality conditions:

$$\left(\mathrm{D}^T\mathrm{D} + \beta_r \mathrm{I}\right) t = o - r^{(k)} + \frac{\lambda_r^{(k)}}{\beta_r}. \tag{57}$$

Subproblem (54) is a quadratically constrained quadratic program solved via the function `fmincon()` performing constrained nonlinear multivariable optimization, included in the Matlab optimization package. □

8 Numerical Results

In this section, we validate the performance of the proposed PC-optimizer framework employed for the signal decomposition task. In detail, the results of Algorithm 1 are illustrated in Sect. 8.1, and comparisons with other state-of-the-art variational decomposition algorithms are given in Sect. 8.2. We finally propose in Sect. 8.3 a more realistic setup where the composed signal f is a mix of four components, including both white additive Gaussian noise and high oscillations.

The experimental setup for Algorithm 1 is described in the following.

The `Estimate_param()` procedure - step [a], Algorithm 1 - evaluates $\mathcal{J}(x; a = 0, \gamma_1, \gamma_2)$ on a coarse parameter grid $[\underline{\gamma}_1 = 10^{-3}, \overline{\gamma}_1 = 10^3] \times [\underline{\gamma}_2 = 10^{-4}, \overline{\gamma}_2 = 10^4]$ for a total of $N_p = 100$ values. The ADMM-based algorithm for each pair (γ_1, γ_2) is stopped as soon as either 3000 iterations are reached or the relative change defined as

$$\delta(x^{(k)}) = \frac{\|x^{(k)} - x^{(k-1)}\|_2}{\|x^{(k-1)}\|_2}, \tag{58}$$

drops below 10^{-9}. Then the estimate of the parameters $(\widehat{\gamma}_1, \widehat{\gamma}_2)$ is obtained by solving the minimization problem (45).

The `Predictor`$(\mathcal{J}(x; a = 0, \widehat{\gamma}_1, \widehat{\gamma}_2))$ solves (7) by the ADMM-based approach up to 5000 iterations.

The `Estimate_param()` procedure - step [c], Algorithm 1 - evaluates $\mathcal{J}(y; a, \gamma_1, \gamma_2)$ on a refined parameter grid $[\underline{\gamma}_1 = \widehat{\gamma}_1, \overline{\gamma}_1 = 200\widehat{\gamma}_1] \times [\underline{\gamma}_2 = \widehat{\gamma}_2/200, \overline{\gamma}_2 = 200\widehat{\gamma}_2]$ for a total of $N_p = 2500$ values. For each pair (γ_1, γ_2) in the grid, the ADMM-based algorithm is stopped as soon as either 10,000 iterations are reached or the relative change $\delta(x^{(k)})$ drops below 10^{-9}. The parameter a is estimated as the minimum discontinuity jump in the cartoon component $\hat{c}$ of the signal result from the `Predictor` step, i.e.

$$a = 1/\min_{|D\hat{c}|\neq 0}(|D\hat{c}|). \tag{59}$$

The value for the parameter β must satisfy condition (36) and thus is set to be $\beta = 5(4a\gamma_1)$. Then the estimate of the parameters $(\widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2)$ is obtained by solving the minimization problem (45).

Finally, the `Corrector`$(\mathcal{J}(\widehat{y}; a, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2)$ solves (7) by the ADMM-based approach up to 10,000 iterations.

The quantitative evaluation of the decomposition results is provided by means of joint signal-to-noise ratio (SNR) over component vector variable $y = (c; s; o)$, defined as

$$\mathrm{SNR}(y) = 10\log_{10}\left(\frac{\|y_{GT} - mean(y_{GT})\|_2^2}{\|y_{GT} - y\|_2^2}\right), \tag{60}$$

with y_{GT} being the reference ground truth component vector.

8.1 Performance of the PC-Optimizer

The performance of the proposed PC-optimizer outlined in Algorithm 1 is here evaluated by two synthetic 1D signal data $f_1 \in \mathbb{R}^{1024}$ and $f_2 \in \mathbb{R}^{512}$ illustrated in the first row of Figs. 2 and 3, respectively. The composed signals consist of piece-wise constant c_{GT}, smooth s_{GT} and oscillatory o_{GT} components reported in the first and the second rows of Figs. 2, 3. We notice that the component o_{GT} of f_1 represents white Gaussian noise, whereas o_{GT} of f_2 contains structured oscillations. For the decomposition of

both signals f_1 and f_2, we run the PC-optimizer in Algorithm 1. Results are reported from the third row of Fig. 2, and 3, column-wise: (left) step [a] and step [b], (right) step [c] and step [d].

For the composed signal f_1 in Fig. 2 the MPCCP values $C(\gamma_1, \gamma_2)$ evaluated in the parameter grid $[\underline{\gamma}_1, \overline{\gamma}_1] \times [\underline{\gamma}_2, \overline{\gamma}_2]$ are shown as a false-colored map in Fig. 2, third row. `Estimate_param()` selects a minimum at $(\widehat{\gamma}_1, \widehat{\gamma}_2) = (0.0562, 1)$, attaining $C(\widehat{\gamma}_1, \widehat{\gamma}_2) = -1.1572$, marked as • in the map $C(\gamma_1, \gamma_2)$. To illustrate quantitative performance of the proposed PC-optimizer alongside $C(\gamma_1, \gamma_2)$, we report the associated SNR grid, with $\max(SNR(y))$ marked by ♦. As expected, the yellow valley-shaped region of small $C(\gamma_1, \gamma_2)$ values, indicating good component separation, corresponds to the area of maximum SNR values. The signal components "c","s" and "o" obtained by applying `Predictor`$(\mathcal{J}(x; a = 0, \widehat{\gamma}_1, \widehat{\gamma}_2))$ are reported in red from the fifth to the seventh row of Fig. 2, superimposed on the corresponding black-colored ground truth signals. Figure 2, right panel, illustrates the MPCCP map produced by step [c] `Estimate_param()`. The minimum $\min(C(\gamma_1, \gamma_2)) = -1.1782$ is attained at $(\widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2) = (0.3615, 13.7830)$, marked as •. Finally, the decomposition results of the `Corrector`$(\mathcal{J}(\widehat{y}, a = 54.6448, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2))$ are reported in the last three rows of the right panel, attaining $SNR(y) = 23.6896$.

In a similar way, in Fig. 3 the decomposition results obtained by applying the PC-optimizer on the synthetic signal f_2 are illustrated. The left panel in Fig. 3 reports the MPCCP map obtained in step [a] `Estimate_param()`. The minimum value $C(\widehat{\gamma}_1, \widehat{\gamma}_2) = -0.9179$ is obtained at $(\widehat{\gamma}_1, \widehat{\gamma}_2) = (0.0316, 0.4642)$, and is marked as • in the reported map $C(\gamma_1, \gamma_2)$. From the fifth to the seventh row of Fig. 3, we report the signal components "c","s" and "o" achieved by the `Predictor`$(\mathcal{J}(x; a = 0, \widehat{\gamma}_1, \widehat{\gamma}_2))$ in red, superimposed on the corresponding black-colored ground truth signals. In the right panel of Fig. 3 we report the results of step [c], namely `Estimate_param()` which found $\min(C(\gamma_1, \gamma_2)) = -1.0193$ at $(\widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2) = (0.6141, 3.4760)$, and step [d], `Corrector`$(\mathcal{J}(\widehat{y}; a = 38.46, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2))$ whose resulting decomposition y is reported in red in the remainder of the column attening $SNR(y) = 21.7992$.

Regions characterized by small values of $C(\gamma_1, \gamma_2)$ again matches regions with high SNR values, thus confirming the goodness of the MPCCP criterium applied to discriminate between components.

8.2 Comparison with Other Signal Decomposition methods

In this section, we validate and explore the potential of the proposed variational optimization model (3)–(4) solved by the PC-optimizer strategy by comparing the decomposition results obtained with other state-of-the-art methods. In particular we selected JOT [8] via its public implementation [9], MB variational approach [30], which is available to public at https://github.com/cvxgrp/signal-decomposition, and GNC, via a custom implementation.

At this aim, we generated synthetic 1D signals $f_3 \in \mathbb{R}^{400}$ and $f_4 \in \mathbb{R}^{500}$ reported together with their components in the upper panels of Figs. 4 and 5, respectively. The model parameters required in all the three algorithms have been tuned to obtain the best achievable decomposition y. In case of MB method, we selected piecewise-constant,

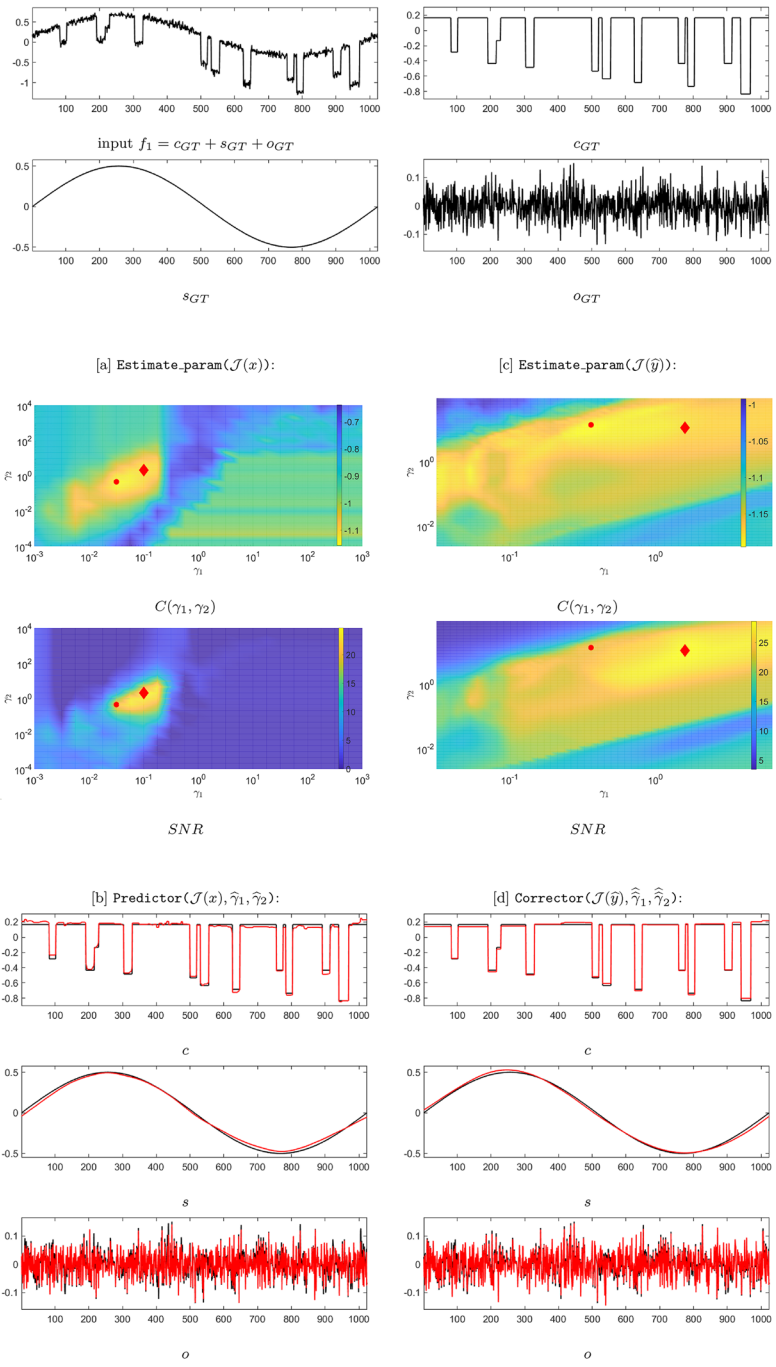

Fig. 2 Synthetic data f_1 and its signal components c_{GT}, s_{GT}, o_{GT} (upper part); decomposition results by the PC-optimizer Algorithm: results of the steps **a**–**c** and **d** (lower part)

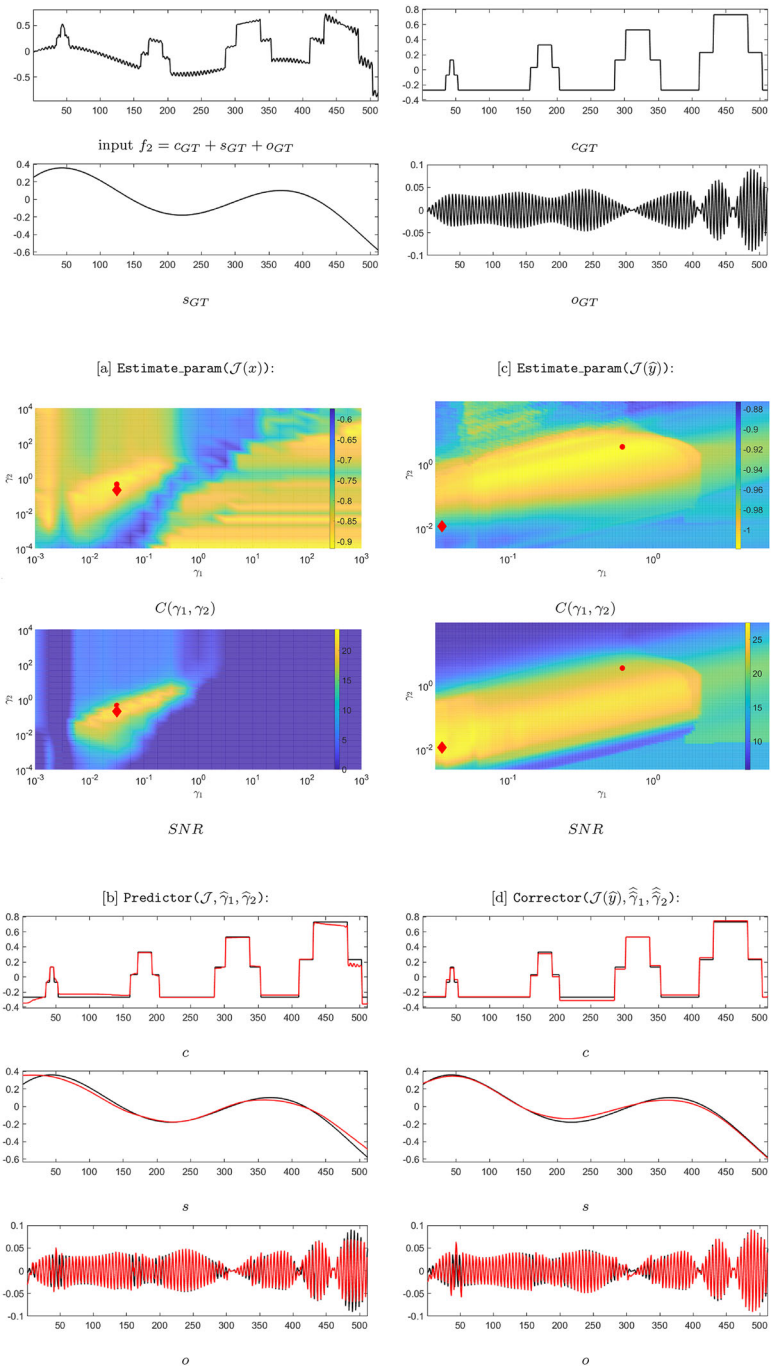

Fig. 3 Synthetic data f_2 and its signal components c_{GT}, s_{GT}, o_{GT} (upper part); decomposition results by the PC-optimizer Algorithm: results of the steps **a**–**c** and **d** (lower part)

smooth and L_2 residual classes to model each component. In the lower panels of Figs. 4 and 5, we report column-wise in red the decomposition result $y = \{c, s, o\}$ of the PC-optimizer, JOT, MB and GNC decompositions, respectively, overlaying with the black-colored ground truth components.

Results of the decomposition of signal f_3 into "c","s" and "o" components are illustrated in Fig. 4, lower panel. The resulting SNR values attained are $SNR(y_{PC}) = 32.42$, $SNR(y_{JOT}) = 26.27$, $SNR(y_{MB}) = 26.36$, $SNR(y_{GNC}) = 18.47$. By a visual inspection we observe that the smooth component "s" is recovered best by the PC-optimizer, while a small interchange of magnitudes between "c" and "o" is present. This is observed also in the components obtained by JOT, MB methods, but in a more pronounced way, probably due to the use of a soft constraint (data fidelity term) in the variational models. MB uses the residual from fidelity term to capture the last component "o" of the decomposition, while JOT model contains a dedicated term for the oscillatory component and uses the fidelity term as a soft constraint for $f = c + s + o$.

In case of the signal f_4, its distinctiveness from signal f_3 lies in the c component, which is binary piece-wise constant. This synthetic signal has been generated by the random generator outlined in [Sec. 2.9] [30] to reproduce a similar example showcased by the authors [Fig. 2.1] [30]. To follow their example, we have selected the boolean, smooth and L_2 residual classes for the decomposition. The authors of JOT model [8] presented a constrained version of the model for binary components as well, therefore, we adapted the $\texttt{Corrector}(\mathcal{J}(\widehat{y}; a, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2))$ procedure to include a boolean constraint. The decomposition results for signal f_4 are reported column-wise in Fig. 5, lower panel. The resulting joint SNR values are $SNR(y_{PC}) = 28.95$, $SNR(y_{JOT}) = 28.73$, $SNR(y_{MB}) = 20.81$, $SNR(y_{GNC}) = 16.32$. From a qualitative inspection of the results, due to the relative closeness between the proposed variational model (3)–(4) and the one presented in JOT approach, the results are quite alike. The PC-optimizer generated artifacts in the last samples of the c component, while JOT results lack in recovery of the s component. The MB method misidentified a few noise oscillations as part of boolean discontinuities in "c". As a benefit of the doubt this may have been caused by the parameter selection, however, using a G-norm type regularization for capturing highly oscillating signals such as Gaussian noise have been proven advantageous over the L_2 residual. The low-quality recovery of the "c" component achieved by applying the GNC algorithm for the decomposition of both f_3 and f_4 is due to the lack of a criterium over the increasing of the parameter a, which affects the concavity of the model.

For what concern the execution times of the compared methods, we should say that the numerical methods vary in few aspects, that influence directly the execution times. Moreover, we used a non-optimized MATLAB implementation of the methods. Experimentally the average time for the solution of the decomposition model by the proposed PC-Optimizer and the compared methods [8] and [30], takes up to 10–20 s for a signal of length up to N=1024, GNC is computationally more expensive. Finally, the proposed PC-Optimizer could be implemented totally in parallel and thus the timings can be further reduced.

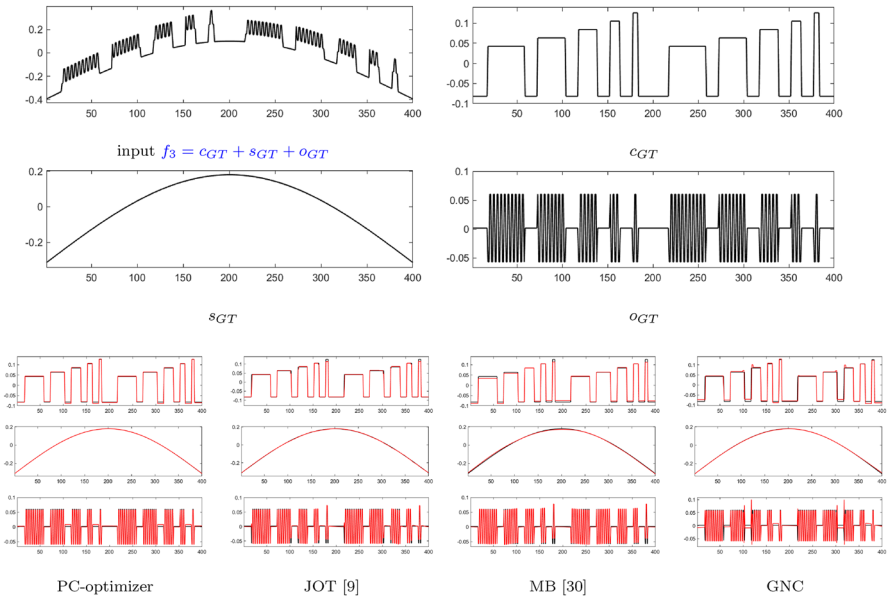

Fig. 4 Synthetic data f_3 and its signal components c_{GT}, s_{GT}, o_{GT} (upper part) and comparison of signal f_3 decomposition results into c, s, o components (lower part). From left to right: PC-Optimzer $SNR(y_{PC}) = 32.42$, JOT method $SNR(y_{JOT}) = 26.27$, MB method $SNR(y_{MB}) = 26.36$ and GNC algorithm $SNR(y_{GNC}) = 18.47$

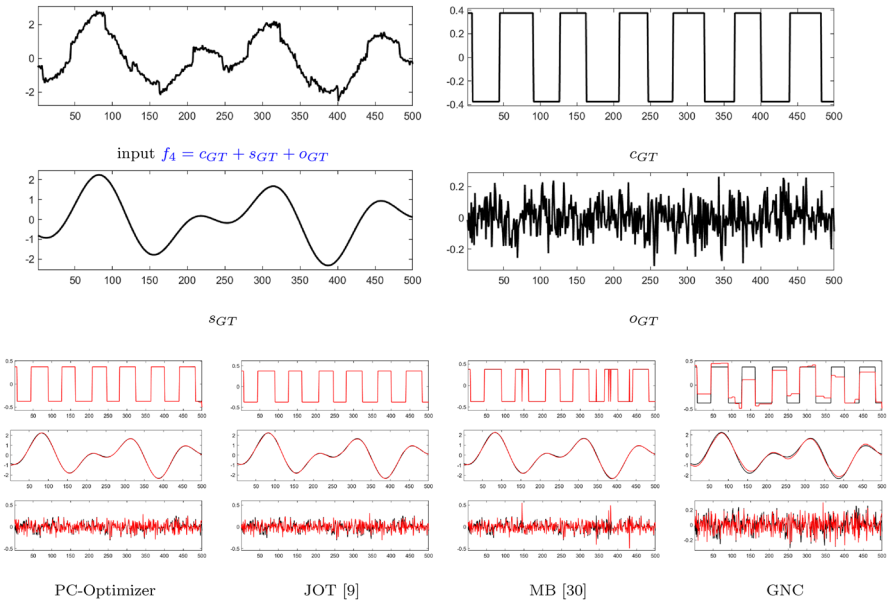

Fig. 5 Synthetic data f_4 and its signal components c_{GT}, s_{GT}, o_{GT} (upper part) comparison of signal f_4 decomposition results into c, s, o components (lower part). From left to right: PC-Optimizer $SNR(y_{PC}) = 28.95$, JOT method $SNR(y_{JOT}) = 28.73$, MB method $SNR(y_{MB}) = 20.81$ and GNC algorithm $SNR(y_{GNC}) = 16.32$

8.3 Quaternary Decomposition $f = c + s + n + t$

In the previous examples, illustrated in Sect. 8.1 and Sect. 8.2, the synthetically generated signals were free of additional corruption component and thus could be decomposed cleanly into oscillating part, trend (piece-wise smooth) and cartoon components. For an arbitrary real signal this assumption frequently fails, as the signal may be corrupted by additive noise which is naturally captured by the oscillatory component. In order to deal with the decomposition of perturbed signals, we propose a post-processing optimization strategy, details of which are provided in Sect. 7, with the aim to split the oscillatory component $o = t + n$, result of the PC-optimizer, into high-frequency oscillations t and additive white noise n. The latter can be a realization of different types among uniform (AWUN), Gaussian (AWGN) and Laplacian (AWLN) noise distributions.

By the way of illustration, we perform the decomposition of a synthetic signal f_5, shown in Fig. 6, upper part, together with the ground truth components: c_{GT}, s_{GT}, t_{GT}, synthetically corrupted by white Gaussian noise $n_{GT} \in \mathcal{N}(0, \sigma)$, with standard deviation $\sigma = 0.15$. As reported in the middle part of Fig. 6, both $\texttt{Predictor}(\mathcal{J}(x), \widehat{\gamma}_1, \widehat{\gamma}_2)$—left panel—and $\texttt{Corrector}(\mathcal{J}(\widehat{y}), \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2)$—right panel—result in one oscillatory component $o = t + n$ which incorporate the noise. Each resulting component is red-colored, while the ground truth components are black-colored. The noise is finally isolated by solving the decomposition problem (52); in the lower part of Fig. 6 the two separated components t and n are shown.

In addition to the decomposition results of the signal f_5 we report the empirical convergence behaviour of the PC-optimizer algorithm. In Fig. 7 we give empirical evidence on the convergence of the sequence generated by the iterative $\texttt{Predictor}(\mathcal{J}(x, a = 0, \widehat{\gamma}_1, \widehat{\gamma}_2))$ in solid red and $\texttt{Corrector}(\mathcal{J}(\widehat{y}, a, \widehat{\widehat{\gamma}}_1, \widehat{\widehat{\gamma}}_2))$ in dashed blue. In particular, the plots report the decreasing of the objective function $\mathcal{J}$ (left), the relative errors $err_k^{(GT)}$ (middle), and the increasing of SNR values (right) along the first 5000 ADMM k iterations. The plots confirm the fast convergence of both the optimization stages of the PC-optimizer for the decomposition of signal f_5; quite similar results are obtained for the other tested signals.

Finally, to further motivate the benefit of the proposed Predictor–Corrector strategy, we compared the results of applying the $\texttt{Corrector}(\cdot)$ alone, starting the ADMM iterations with different initial guesses. In particular, in Fig. 8, each column represents a different initialization to $\texttt{Corrector}(\cdot)$ from left to right: $x^{(0)}$ is the result of the $\texttt{Predictor}(\cdot)$, $x^{(0)}$ is the zero vector, $x^{(0)} = (c^{(0)}, g^{(0)}) = (0_N, f)$, $x^{(0)}$ is a zero-mean random vector with Gaussian distribution of standard deviation equal to the one of the ground truth components c, $\sigma_c = 0.0202$, and g, $\sigma_g = 0.0248$. Each row represents the SNR values for a parameter grid $[\underline{\gamma}_1, \overline{\gamma}_1] \times [\underline{\gamma}_2, \overline{\gamma}_2]$ after 1000 (first row) and 60,000 (second row) ADMM iterations, the maximum SNR value $\max(SNR(x))$ is marked by ♦. As expected, different initial guesses for the $\texttt{Corrector}(\cdot)$ lead to different local minimizers. Initializing the $\texttt{Corrector}(\cdot)$ with the result $\widehat{y}$ from the $\texttt{Predictor}(\cdot)$ leads to a more efficient convergence into an attraction basin of a better quality minimizer.

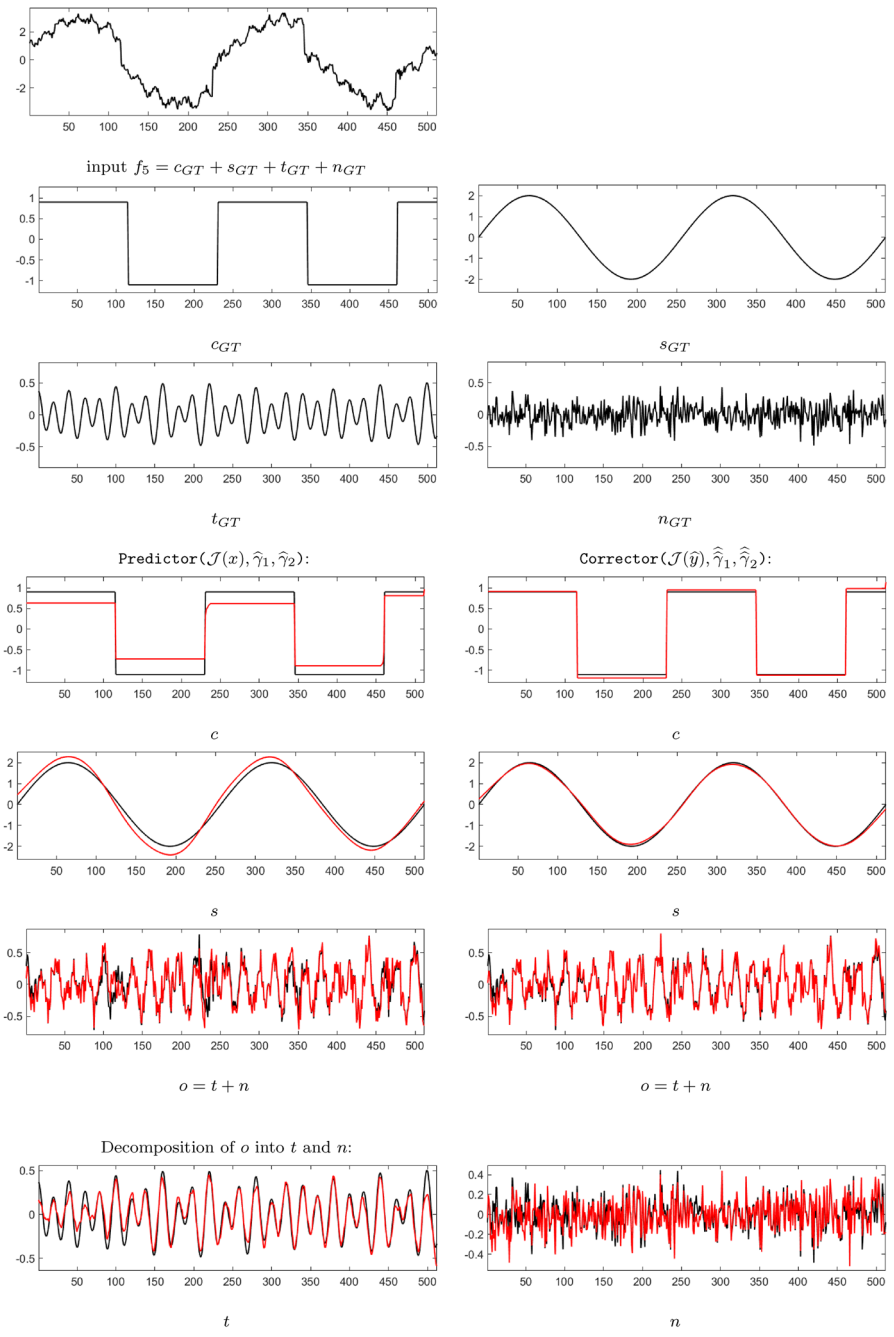

Fig. 6 Synthetic data f_5 and its signal components $c_{GT}, s_{GT}, t_{GT}, n_{GT}$ (upper part) and its decomposition results: PC-optimizer (middle part) and denoiser (bottom part)

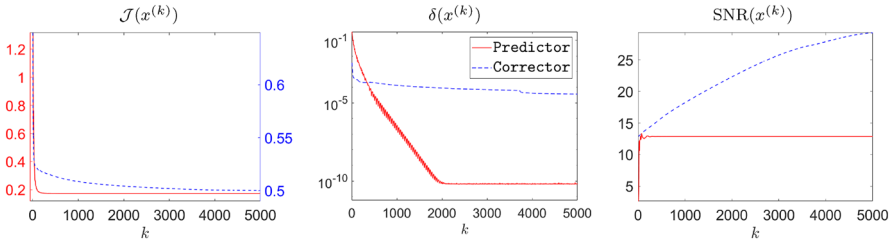

Fig. 7 Empirical convergence of the ADMM algorithm applied for solving `Predictor`$(\cdot)$ and `Corrector`$(\cdot)$ models for the decomposition of signal f_5. From left to right: plot of the $\mathcal{J}$ as function of the number of iterations k, relative error change on the ADMM solution (see (58)), SNR values (see (60)) over the component vector $y = (c; s; o)$

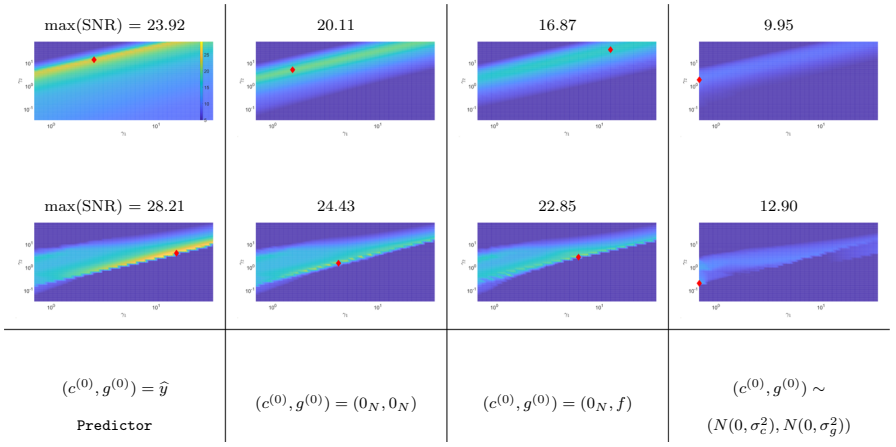

Fig. 8 SNR grid values for the decomposition of signal f_5 using the `Corrector`$(\cdot)$ alone with different initial guesses; from top to bottom after 1000 and 60,000 ADMM iterations. Each column represents a different initialization to `Corrector`$(\cdot)$ (reported in the last row). The maximum achieved SNR value is reported and marked by ♦ in the grid

9 Conclusions

We proposed the PC-optimizer framework, a two stage algorithmic approach inspired by the Predictor–Corrector strategy to efficiently address the challenging minimization of a multi-parameter functional with a nonconvex parametric penalty. This minimization problem is quite common in signal/image processing applications to enforce sparsity in the space of solutions or in its transforms. The two main challenges that the PC-optimizer faces are the computation of optimal multi-parameters of the functional, and the choice of a good initial guess that can allow a fast algorithmic convergence to a good local minimizer. The benefits of the PC-optimizer have been highlighted for the solution of a decomposition variational model aimed at the ternary decomposition of 1D signals into cartoon, smooth and oscillatory components. A model analysis is provided and we presented an ADMM-based algorithm for the minimization of both Predictor and Corrector stages which allows to guarantee the convergence of

the entire PC-optimizer framework. Numerical experiments have been carried out and the reported results strongly indicate that the PC-optimizer applied to the solution of the variational decomposition model is able to achieve excellent performance while remaining in a nonconvex regime. This paves the way for a facilitated solution of an entire class of nonconvex optimization problems.

Appendix

Proof of Proposition 4.1

Proof It comes immediately from definitions of the cost function $\mathcal{J}$ in (8) and of the penalty function ϕ in (5) that $\mathcal{J}$ is proper, continuous and bounded below by zero in $(c, g)^T$, for any $f \in \mathbb{R}^N$, $\gamma_1, \gamma_2 \in \mathbb{R}_{++}$, $a \in \mathbb{R}_+$. Then, proving that $\mathcal{J}$ is constant along straight lines of direction defined by the vector d in (15)—which clearly implies non-coercivity of $\mathcal{J}$—is straightforward. In fact, after noting that such a line bundle can be defined parametrically as $(c; g) + t\, d = (c + t\, 1_N\,; g)$, with $t \in \mathbb{R}$, we have

$$\begin{aligned}\mathcal{J}(c + t\,1_N, g; a, \gamma_1, \gamma_2) = \gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}c + \cancel{t\,\mathrm{D}1_N})_i\right|; a\right) + \frac{\gamma_2}{2}\|\mathrm{H}(c + \mathrm{D}^{\mathrm{T}} g - f) + \cancel{t\,\mathrm{H}1_N}\|_2^2 \\ + \|g\|_\infty = \mathcal{J}(c, g; a, \gamma_1, \gamma_2) \quad \forall (c; g) \in \mathbb{R}^{2N},\ \forall t \in \mathbb{R}\,, \end{aligned} \tag{61}$$

where we used the fact that constant vectors belong to the null spaces of both D and H, for any boundary conditions—see (12). This implies that, if there exists a minimizer $(\widehat{c}, \widehat{g})^T \in \mathbb{R}^{2N}$ of $\mathcal{J}$ (this will be proved later), the straight line $(\widehat{c}; \widehat{g}) + t\, d$, $t \in \mathbb{R}$, contains an infinity of minimizers equivalent to $(\widehat{c}; \widehat{g})$ in terms of function value.

For what concerns convexity, first we note that when $a = 0$ the function $\mathcal{J}$ in (8) is given by the sum of three convex functions (in fact, the first term in $\mathcal{J}$ becomes the standard, discrete TV semi-norm of c), hence $\mathcal{J}$ is convex in x. For $a > 0$, the first term in $\mathcal{J}$ is nonconvex, hence we start analyzing convexity of $\mathcal{J}$ separately with respect to the optimization variables c and g. To this aim, we introduce the two functions

$$\mathcal{J}_g(g; a, \gamma_1, \gamma_2) := \mathcal{J}(\bar{c}, g; a, \gamma_1, \gamma_2) = \frac{\gamma_2}{2}\left\|\mathrm{H}\mathrm{D}^{\mathrm{T}} g - \bar{v}_1\right\|_2^2 + \|g\|_\infty + \bar{v}_2, \tag{62}$$

$$\begin{aligned}\mathcal{J}_c(c; a, \gamma_1, \gamma_2) := \mathcal{J}(c, \bar{g}; a, \gamma_1, \gamma_2) = \frac{\gamma_2}{2}\|\mathrm{H}c - \bar{w}_1\|_2^2 \\ + \gamma_1 \sum_{i=1}^{N} \phi\left(|(\mathrm{D}c)_i|\,; a\right) + \bar{w}_2\,, \end{aligned} \tag{63}$$

where $\bar{c} \in \mathbb{R}^N$ in (62) and $\bar{g} \in \mathbb{R}^N$ in (63) are any two fixed values of variables c and g, respectively, and where $\bar{v}_1 \in \mathbb{R}^N$, $\bar{v}_2 \in \mathbb{R}_+$ in (62) are constants with respect to g whereas $\bar{w}_1 \in \mathbb{R}^N$, $\bar{w}_2 \in \mathbb{R}_+$ in (63) are constants with respect to c. Function $\mathcal{J}_g$ in (62) is clearly convex and coercive in g, as it is the sum of a constant and two non-negative convex functions, the latter ($\|g\|_\infty$) being coercive. However, it is not strictly convex as matrix $\mathrm{H}\mathrm{D}^{\mathrm{T}}$ has not full rank. Then, in order to prove that function $\mathcal{J}_c$ in (63)

is not convex in c—which clearly implies that the total function $\mathcal{J}$ in (8) is nonconvex in x—we consider its restriction to a line in its domain $c(t) = t\,\bar{z}, t \in \mathbb{R}$, with generic direction $\bar{z} \in \mathbb{R}^N$. After some simple algebraic manipulations, the restriction reads

$$\mathcal{J}_c(c(t); a, \gamma_1, \gamma_2) = \frac{\gamma_2}{2} \sum_{i=1}^{N} \left(\bar{p}_i t - \bar{w}_{1,i}\right)^2 + \gamma_1 \sum_{i=1}^{N} \phi\left(|\bar{q}_i|\,|t|; a\right), \quad \bar{p} = \mathrm{H}\bar{z}, \;\; \bar{q} = \mathrm{D}\bar{z}. \tag{64}$$

If anti-reflective BC are adopted for D and H, taking $\bar{z} = l_N$ yields $\bar{p} = 0_N$ and $\bar{q} \neq 0_N$—see (12)—hence the restriction in (64) reduces to the sum of a constant and a nonconvex function of t. It follows that, for anti-reflective BC, the total function $\mathcal{J}$ in (8) is nonconvex in x for any $f \in \mathbb{R}^N$ and any $\gamma_1, \gamma_2, a \in \mathbb{R}_{++}$. In case of periodic or reflective BC, for the same direction $\bar{z} = l_N$ we have $\bar{p}, \bar{q} \neq 0_N$, so the restriction in (64) is the sum of a quadratic convex function and a nonconvex function of t. This implies that there exist values of parameters γ_1, γ_2, a yielding nonconvexity of the restriction and, hence, of the total $\mathcal{J}$.

We finally prove that, in spite of its non-coercivity and possible nonconvexity, $\mathcal{J}$ in (8) indeed admits (an infinity of) global minimizers for any $f \in \mathbb{R}^N$ and any $\gamma_1, \gamma_2 \in \mathbb{R}_{++}$, $a \in \mathbb{R}_+$. As $\mathcal{J}$ is continuous and bounded below by zero, we proceed as follows. First, we detect all the possible paths of non-coercivity for $\mathcal{J}$, namely paths towards infinity in the domain $\mathbb{R}^{2N}$ of $\mathcal{J}$ along which the value of $\mathcal{J}$ does not tend to $+\infty$. Then, we compute all the possible limit values of $\mathcal{J}$ along these paths and, finally, we prove that these or lower values are attained at (finite) domain points $x \in \mathbb{R}^{2N}$. This implies that, even if a global infimizer exists at infinity, then a corresponding—i.e., characterized by the same function value—global minimizer exists as well.

Paths of non-coercivity. Function $\mathcal{J}$ in (8) is clearly coercive in g due to the coercivity in g of the last term $\|g\|_\infty$. It follows that possible paths of non-coercivity for $\mathcal{J}$ must be sought by keeping g bounded, that is, letting only $\|c\|_2$ approach $+\infty$. Then, the sum of the first and second terms of $\mathcal{J}$ in (8),

$$\gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}c)_i\right|; a\right) + \frac{\gamma_2}{2} \left\|\mathrm{H}c - \mathrm{H}\left(f - \mathrm{D}^{\mathrm{T}} g\right)\right\|_2^2, \tag{65}$$

is always non-coercive in c, but the paths of non-coercivity in c depend both on the value of parameter a and on the BC adopted for linear operators D and H. The paths of non-coercivity in c for the quadratic term in (65) are clearly only those approaching at infinity a direction parallel to null(H) which, we recall, contains constant vectors for periodic and reflective BC, constant and linear (i.e., affine) vectors for anti-reflective BC—see (12). Then, due to the properties of function ϕ in (5), the first term in (65) is bounded from below (by zero) and, for $a > 0$, also from above (by $\gamma_1 N < +\infty$). Hence, for $a > 0$ the term does not affect the coercivity of function $\mathcal{J}$. For $a = 0$, the first term reduces to the standard, discrete TV semi-norm of c, which is coercive except clearly along paths approaching at infinity a direction parallel to null(D), the

set of constant vectors. It follows from all the previous considerations that the paths of non-coercivity in $x = (c, g)^T$ for the total function $\mathcal{J}$ in (8) are those approaching at infinity a direction parallel to the linear subspace $\mathcal{S} \subset \mathbb{R}^{2N}$ given by

$$\mathcal{S} = \begin{cases} \text{span}\big(d\big) & \text{for } a = 0,\ \text{BC} \in \{\text{P, R, A}\}, \\ \text{span}\big(d\big) & \text{for } a > 0,\ \text{BC} \in \{\text{P, R}\}, \\ \text{span}\big(d,\ (l_N; 0_N)\big) & \text{for } a > 0,\ \text{BC} = \text{A}, \end{cases} \tag{66}$$

with vector $d = (1_N, 0_N)^T$ introduced in (15).

Limit values. We note that in the first two cases of (66) the paths of non-coercivity for $\mathcal{J}$ are only those with asymptotic direction given by vector d, along which $\mathcal{J}$ is constant. It clearly follows that any limit value achieved by $\mathcal{J}$ at infinity along these paths is also achieved for some finite $x = (c, g)^T \in \mathbb{R}^{2N}$. Then, in order to compute the limit values of $\mathcal{J}$ along its paths of non-coercivity for the last case in (66)—which corresponds to $a > 0$ and anti-reflective BC for D and H - it suffices to analyze the behaviour at infinity of the restrictions $\mathcal{J}^{(\bar{c},\bar{g})}$ of $\mathcal{J}$ to the family of parameterized affine subspaces $\mathcal{S}^{(\bar{c},\bar{g})} \subset \mathbb{R}^{2N}$, with parameter $(\bar{c}; \bar{g}) \in \mathbb{R}^{2N}$, of the form

$$\mathcal{S}^{(\bar{c},\bar{g})} \ = \ (\bar{c}; \bar{g}) + \mathcal{S}, \tag{67}$$

with $\mathcal{S}$ defined in (66). Based on (66)–(67) and on the definition of $\mathcal{J}$ in (8), using that $\text{D}1_N = \text{H}1_N = 0_N$ for any BC, and dropping (for shortness of notation) the dependence of $\mathcal{J}$ and its restrictions $\mathcal{J}^{(\bar{c},\bar{g})}$ on the parameters a, γ_1, γ_2, after simple manipulations the restrictions read

$$\begin{aligned} \mathcal{J}^{(\bar{c},\bar{g})}(t_1, t_2) \ &= \ \mathcal{J}\,(\bar{c} + t_1 1_N + t_2\, l_N,\ \bar{g}) \\ &= \ \overbrace{\gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\text{D}\,(\bar{c} + t_2\, l_N))_i\right|; a\right)}^{G^{(\bar{c},\bar{g})}(t_2)} \\ &\quad \underbrace{+ \frac{\gamma_2}{2}\left\|\text{H}\left(\bar{c} + \text{D}^{\text{T}}\bar{g} - f\right)\right\|_2^2 + \|\bar{g}\|_\infty}_{\overline{G}^{(\bar{c},\bar{g})}}, \qquad (t_1, t_2) \in \mathbb{R}^2, \end{aligned} \tag{68}$$

where the latter term $\overline{G}^{(\bar{c},\bar{g})} \in \mathbb{R}_+$ depends on $(\bar{c}, \bar{g})$ but not on (t_1, t_2). Hence, the behaviour of $\mathcal{J}^{(\bar{c},\bar{g})}(t_1, t_2)$ at infinity—i.e., for $\|(t_1; t_2)\|_2 \to \infty$—depends mainly on the former term $G^{(\bar{c},\bar{g})}(t_2)$ which, in its turn, only depends on t_2. It comes from the definition of matrix D for anti-reflective BC—matrix D_A in (11) - and of vector l_N that $\text{D}\,l_N = 1_N$, hence

$$\left|(\text{D}\,(\bar{c} + t_2\, l_N))_i\right| \ = \ \left|(\text{D}\bar{c})_i + t_2\right| \quad \forall\, i = 1, \ldots, N\,. \tag{69}$$

Taking the limit for $\|(t_1; t_2)\|_2$ tending to $+\infty$, with a little abuse of notation (the limit does not formally exist), we have

$$\lim_{\|(t_1;t_2)\|_2\to\infty} \left|(\mathrm{D}\bar{c})_i + t_2\right| = \begin{cases} +\infty & \text{if } |t_2| \to +\infty, \\ \left|(\mathrm{D}\bar{c})_i + \bar{t}_2\right| < +\infty & \text{if } |t_1| \to +\infty,\ t_2 \to \bar{t}_2 \in \mathbb{R}, \end{cases} \tag{70}$$

$i = 1, \ldots, N$. Based on (68)–(70) and on property $\lim_{t\to+\infty} \phi(t; a) = \frac{1}{2a}$ for $a > 0$, we can thus write

$$\lim_{\|(t_1;t_2)\|_2\to\infty} \mathcal{J}^{(\bar{c},\bar{g})}(t_1, t_2) = \begin{cases} L_1^{(\bar{c},\bar{g})} & \text{if } |t_2| \to +\infty, \\ L_2^{(\bar{c},\bar{g})} & \text{if } |t_1| \to +\infty,\ t_2 \to \bar{t}_2 \in \mathbb{R}, \end{cases} \tag{71}$$

with limit values $L_1^{(\bar{c},\bar{g})}, L_2^{(\bar{c},\bar{g})} \in \mathbb{R}_+$ given by

$$L_1^{(\bar{c},\bar{g})} = \gamma_1 \frac{N}{2a} + \overline{G}^{(\bar{c},\bar{g})}, \qquad L_2^{(\bar{c},\bar{g})} = \gamma_1 \sum_{i=1}^{N} \phi\left(\left|(\mathrm{D}\bar{c})_i + \bar{t}_2\right|; a\right) + \overline{G}^{(\bar{c},\bar{g})}. \tag{72}$$

After noting that $L_2^{(\bar{c},\bar{g})} \le L_1^{(\bar{c},\bar{g})}$, we complete the proof by demonstrating that, for any $f \in \mathbb{R}^N$, $\gamma_1, \gamma_2, a \in \mathbb{R}_{++}$, $(\bar{c}, \bar{g})^T \in \mathbb{R}^{2N}$ and any $\bar{t}_2 \in \mathbb{R}$, there exists a point $x = (c, g)^T \in \mathbb{R}^{2N}$ (not at infinity) such that $\mathcal{J}(c, g; a, \gamma_1, \gamma_2) \le L_2^{(\bar{c},\bar{g})}$. In fact, e.g., for

$$x = (c, g)^T = \left(\bar{c} + \bar{t}_2 l_N,\ \bar{g}\right)^T, \tag{73}$$

we have that $\mathcal{J}(c, g; a, \gamma_1, \gamma_2) = L_2^{(\bar{c},\bar{g})}$. □

Proof of Proposition 4.2

Proof We first recall that any hyperplane $\mathcal{H}_v$ defined as in (17) is not parallel to the vector d in (15) (due to condition $v^{\mathrm{T}} d \neq 0$), hence it intersects any straight line of direction d in one and only one point. Since according to Proposition 4.1 the function $\mathcal{J}$ admits at least a line with direction d of equivalent global minimizers, the restriction of $\mathcal{J}$ to any feasible set $\mathcal{H}_v$ also admits global minimizers which are characterized by the same (minimum) cost function value. However, the dimensionality of the set of global minimizers is reduced by 1. □

Proof of Corollary 4.1

Proof In the proof of Proposition 4.2 we demonstrated that the paths of non-coercivity in the total optimization variable $x = (c, g)^T$ for function $\mathcal{J}$ defined in (8) are only those approaching at infinity a direction parallel to the linear subspace $\mathcal{S} \in \mathbb{R}^{2N}$ given in (66), depending on parameter a and on the BC for finite difference operators D and H. It follows easily that the paths of non-coercivity for $\mathcal{J}$ constrained to any

hyperplanar set $\mathcal{H}_v$ defined as in (17) are those with asymptotic direction parallel to the linear subspace $\widetilde{\mathcal{S}}$ given by the intersection of $\mathcal{S}$ and $\mathcal{H}_v$. In formula, we have that

$$\widetilde{\mathcal{S}} := \mathcal{S} \cap \mathcal{H}_v = \begin{cases} \{0_{2N}\} & \text{for } a = 0,\ \text{BC} \in \{\text{P}, \text{R}, \text{A}\}, \\ \{0_{2N}\} & \text{for } a > 0,\ \text{BC} \in \{\text{P}, \text{R}\}, \\ \text{span}\big((l_N, 0_N)^T \big) & \text{for } a > 0,\ \text{BC} = \text{A}. \end{cases} \tag{74}$$

Note that the above intersection of $\mathcal{S}$ with $\mathcal{H}_v$ which, we recall, contains homogeneous hyperplanes not parallel to vector d, simply removes d from the spanned directions in (66). It follows from (74) that $\mathcal{J}$ constrained to any set $\mathcal{H}_v$ is coercive in $x = (c, g)^T$ when $a = 0$ or when $a > 0$ and $\text{BC} \in \{\text{P}, \text{R}\}$, whereas it has the direction $(l_N, 0_N)$ of asymptotic non-coercivity when $a > 0$ and $\text{BC} = \text{A}$. □

Proof of Lemma 4.1

Proof It follows from definition (18) that

$$\begin{aligned} \text{null}\,(\text{Q}_{\text{BC}}) &= \left\{ x = (c, g)^T \in \mathbb{R}^{2N} : \quad \text{H}_{\text{BC}}\,[\,\text{I}_N\,,\ \text{D}_{\text{BC}}^{\text{T}}\,]\,x \ = \ 0_N \right\} \\ &= \left\{ x = (c, g)^T \in \mathbb{R}^{2N} : \quad c + \text{D}_{\text{BC}}^{\text{T}}\, g \ \in \ \text{null}(\text{H}_{\text{BC}}) \right\}. \end{aligned} \tag{75}$$

Hence, by recalling the expressions of $\text{null}(\text{H}_{\text{BC}})$ in (12), we have:

$$\text{null}\left(\text{Q}_{\text{P/R}}\right) = \left\{ (c, g)^T \in \mathbb{R}^{2N} : \quad c + \text{D}_{\text{P/R}}^{\text{T}}\, g \ = \ \kappa_1 1_N, \ \ \kappa_1 \in \mathbb{R} \right\}, \tag{76}$$

$$\text{null}\,(\text{Q}_{\text{A}}) = \left\{ (c, g)^T \in \mathbb{R}^{2N} : \quad c + \text{D}_{\text{A}}^{\text{T}}\, g \ = \ \kappa_1 1_N + \kappa_2\, l_N, \ \ \kappa_1, \kappa_2 \in \mathbb{R} \right\}. \tag{77}$$

Projecting the two null spaces above onto the linear span of v_1, adding and subtracting $v_2^{\text{T}} g$, we have

$$v_1^{\text{T}} \left(c + \text{D}_{\text{P/R}}^{\text{T}}\, g \right) + v_2^{\text{T}} g - v_2^{\text{T}} g = v_1^{\text{T}}\, (\kappa_1 1_N)\,, \tag{78}$$

and

$$v_1^{\text{T}} \left(c + \text{D}_{\text{A}}^{\text{T}}\, g \right) + v_2^{\text{T}} g - v_2^{\text{T}} g = v_1^{\text{T}}\, (\kappa_1 1_N + \kappa_2 l_N)\,. \tag{79}$$

Hence, intersecting with the constraint set $\mathcal{H}_v$ defined in (17) (or, equivalently, imposing the constraint $v_1^{\text{T}} c + v_2^{\text{T}} g = 0$) yields, respectively,

$$\kappa_1 = \frac{1}{v^{\text{T}} d} \left(v_1^{\text{T}} \text{D}_{\text{P/R}}^{\text{T}} g - v_2^{\text{T}} g \right), \tag{80}$$

and

$$\kappa_1 = \frac{1}{v^{\mathrm{T}}d}\left(v_1^{\mathrm{T}}\mathrm{D}_{\mathrm{A}}^{\mathrm{T}}g - v_2^{\mathrm{T}}g - \kappa_2(v_1^{\mathrm{T}}l_N)\right). \tag{81}$$

Therefore, replacing (80),(81) in (76),(77) respectively, we have

$$\mathcal{N}_{\mathrm{P/R}} = \left\{(c,g)^T \in \mathbb{R}^{2N} : c = \left(-\mathrm{D}_{\mathrm{P/R}}^{\mathrm{T}} + \frac{1}{v^{\mathrm{T}}d}1_N\left(v_1^{\mathrm{T}}\mathrm{D}_{\mathrm{P/R}}^{\mathrm{T}} - v_2^{\mathrm{T}}\right)\right)g\right\}, \tag{82}$$

$$\mathcal{N}_{\mathrm{A}} = \left\{(c,g)^T \in \mathbb{R}^{2N} : c = \left(-\mathrm{D}_{\mathrm{A}}^{\mathrm{T}} + \frac{1}{v^{\mathrm{T}}d}1_N\left(v_1^{\mathrm{T}}\mathrm{D}_{\mathrm{A}}^{\mathrm{T}} - v_2^{\mathrm{T}}\right)\right)g + \kappa\, h_N,\ \kappa \in \mathbb{R}\right\}, \tag{83}$$

with $h_N := l_N - \left(v_1^{\mathrm{T}}l_N\right)1_N$. Writing g as a linear combination of the canonical basis vectors of $\mathbb{R}^N$,

$$g = \sum_{j=1}^{N}\alpha_j e_N^{(j)}, \tag{84}$$

and then replacing (84) into (82), (83), we have, respectively,

$$\begin{aligned} c &= \left(-\mathrm{D}_{\mathrm{P/R}}^{\mathrm{T}} + \frac{1}{v^{\mathrm{T}}d}1_N\left(v_1^{\mathrm{T}}\mathrm{D}_{\mathrm{P/R}}^{\mathrm{T}} - v_2^{\mathrm{T}}\right)\right)\sum_{j=1}^{N}\alpha_j e_N^{(j)} \\ &= \sum_{j=1}^{N}\alpha_j\left(-\mathrm{D}_{\mathrm{P/R}}^{(j)} + \frac{1}{v^{\mathrm{T}}d}\left(\mathrm{D}_{\mathrm{P/R}}^{(j)}v_1 - v_2^{(j)}\right)1_N^{\mathrm{T}}\right)^{\mathrm{T}}, \end{aligned} \tag{85}$$

and

$$\begin{aligned} c &= \left(-\mathrm{D}_{\mathrm{A}}^{\mathrm{T}} + \frac{1}{v^{\mathrm{T}}d}1_N\left(v_1^{\mathrm{T}}\mathrm{D}_{\mathrm{A}}^{\mathrm{T}} - v_2^{\mathrm{T}}\right)\right)\sum_{j=1}^{N}\alpha_j e_N^{(j)} + \kappa\, h_N \\ &= \sum_{j=1}^{N}\alpha_j\left(-\mathrm{D}_{\mathrm{A}}^{(j)} + \frac{1}{v^{\mathrm{T}}d}\left(\mathrm{D}_{\mathrm{A}}^{(j)}v_1 - v_2^{(j)}\right)1_N^{\mathrm{T}}\right)^{\mathrm{T}} + \alpha_{N+1}h_N, \end{aligned} \tag{86}$$

where in (86) we set $\alpha_{N+1} = \kappa$.

It follows from (85), (86) that the sets in (82), (83) can be respectively written as

$$\begin{pmatrix} c \\ g \end{pmatrix} = \sum_{j=1}^{N}\alpha_j\begin{pmatrix} \left(-\mathrm{D}_{\mathrm{P/R}}^{(j)} + \frac{1}{v^{\mathrm{T}}d}\left(\mathrm{D}_{\mathrm{P/R}}^{(j)}v_1 - v_2^{(j)}\right)1_N^{\mathrm{T}}\right)^{\mathrm{T}} \\ e_N^{(j)} \end{pmatrix}, \tag{87}$$

and

$$\begin{pmatrix} c \\ g \end{pmatrix} = \sum_{j=1}^{N} \alpha_j \begin{pmatrix} \left(-\mathrm{D}_{\mathrm{A}}^{(j)} + \dfrac{1}{v^{\mathrm{T}} d} \left(\mathrm{D}_{\mathrm{A}}^{(j)} v_1 - v_2^{(j)} \right) 1_N^{\mathrm{T}} \right)^{\mathrm{T}} \\ e_N^{(j)} \end{pmatrix} + \alpha_{N+1} \begin{pmatrix} h_N \\ 0_N \end{pmatrix}. \quad (88)$$

This demonstrates statement (19)–(20), in particular the expression of matrix M_{BC}. Finally, it is easy to verify that M_{BC} has full column rank both for $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}\}$ and for $\mathrm{BC} = \mathrm{A}$, which leads to (21) as M_{BC} has N and $N+1$ columns for $\mathrm{BC} \in \{\mathrm{P}, \mathrm{R}\}$ and $\mathrm{BC} = \mathrm{A}$, respectively. □

Proof of Proposition 4.3

Proof It follows straightforwardly from Proposition 4.1 (when $a = 0$, $\mathcal{J}$ is continuous and convex), Corollary 4.1 (when $a = 0$, $\mathcal{J}$ is coercive on any feasible set $\mathcal{H}_v$) and from convexity of the feasible sets $\mathcal{H}_v$ defined in (17), that when $a = 0$ the optimization problem in (16) is convex and admits a m-dimensional convex compact set of solutions $\widehat{x} = (\widehat{c}, \widehat{g})^T \in \mathcal{M} \subset \mathcal{H}_v \subset \mathbb{R}^{2N}$. In particular, $\mathcal{M}$ has dimension $m \leq (2N-1)$, as it must belong to the $(2N-1)$-dimensional hyperplanar feasible set $\mathcal{H}_v$. Moreover, if $\mathcal{M}$ is not a singleton—equivalently, if $m > 0$ and $\mathcal{M}$ is at least a 1-dimensional compact convex set of $\mathcal{H}_v$, that is a bounded and closed segment—then it must belong to a m-dimensional affine subset of $\mathcal{H}_v$, otherwise it would be nonconvex.

In order to obtain more information on the set of solutions $\mathcal{M}$—which, we remark, corresponds to the set of global minimizers of the convex function $\widetilde{\mathcal{J}}$ in (16)—in the following we analyze the restrictions $\widetilde{\mathcal{J}}_\ell$ of $\widetilde{\mathcal{J}}$ to all possible straight lines ℓ belonging to the $(2N-1)$-dimensional hyperplanar feasible set $\mathcal{H}_v$ and check whether these restrictions admit a unique global minimizer or, instead, they can admit a closed and bounded interval of global minimizers. We parametrically define straight lines by

$$\begin{aligned} \ell \subset \mathcal{H}_v: \quad x(t) &= \bar{x} + t\,\bar{z} = \begin{pmatrix} \bar{c} \\ \bar{g} \end{pmatrix} + t \begin{pmatrix} \bar{z}_c \\ \bar{z}_g \end{pmatrix} = \begin{pmatrix} c(t) \\ g(t) \end{pmatrix}, \quad t \in \mathbb{R}, \\ \text{with} \quad \bar{c}, \bar{g}, \bar{z}_c, \bar{z}_g \in \mathbb{R}^N: \quad \bar{x} &= \begin{pmatrix} \bar{c} \\ \bar{g} \end{pmatrix} \in \mathcal{H}_v, \quad \bar{z} = \begin{pmatrix} \bar{z}_c \\ \bar{z}_g \end{pmatrix} \in \mathcal{H}_v. \end{aligned} \quad (89)$$

The restrictions read

$$\begin{aligned} \widetilde{\mathcal{J}}_\ell(t) &:= \widetilde{\mathcal{J}}(c(t), g(t); a = 0, \gamma_1, \gamma_2) \\ &= \gamma_1 \sum_{i=1}^{N} \left| (\mathrm{D}c(t))_i \right| + \frac{\gamma_2}{2} \|\mathrm{Q}\,x(t) - \mathrm{H}f\|_2^2 + \|g(t)\|_\infty + \cancel{\iota_{\mathcal{H}_v}(x(t))}, \end{aligned} \quad (90)$$

with matrix $\mathrm{Q} \in \mathbb{R}^{N \times 2N}$ defined in (18) and where, for shortness of notation, we drop both the subscript BC for matrices Q, D, H and other BC-dependent quantities, and the dependency of restriction $\widetilde{\mathcal{J}}_\ell$ on parameters γ_1, γ_2. Moreover, in (90) we cancel

the indicator function as it is identically zero for any parametric line $x(t) \in \mathcal{H}_v$. After recalling the definition of set $\mathcal{N}$ in (19), introducing the set

$$\mathcal{N}^{\perp} := (\text{null}(\mathrm{Q}_{\mathrm{BC}}))^{\perp} \cap \mathcal{H}_v \,, \tag{91}$$

and indicating with the superscripts $^{\parallel}$ and $^{\perp}$ the projections of a vector in $\mathbb{R}^{2N}$ onto the sets $\mathcal{N}$ in (19) and $\mathcal{N}^{\perp}$ in (91), respectively, we can write any line $x(t) \in \mathcal{H}_v$ defined as in (89) by the direct sum of two orthogonal lines $x^{\parallel}(t)$ and $x^{\perp}(t)$, the former parallel to $\mathcal{N}$, the latter parallel to $\mathcal{N}^{\perp}$. In fact, since according to their definition in (89) we have that $\bar{x}, \bar{z} \in \mathcal{H}_v$, then $\bar{x} = \bar{x}^{\parallel} + \bar{x}^{\perp}$ and $\bar{z} = \bar{z}^{\parallel} + \bar{z}^{\perp}$. Hence, we can write

$$x(t) = \overbrace{\bar{x}^{\parallel} + t\,\bar{z}^{\parallel}}^{x^{\parallel}(t) \in \mathcal{N},\, \forall\, t} + \overbrace{\bar{x}^{\perp} + t\,\bar{z}^{\perp}}^{x^{\perp}(t) \in \mathcal{N}^{\perp},\, \forall\, t} \,. \tag{92}$$

By replacing (92) into (90), we have

$$\begin{aligned}
\widetilde{\mathcal{J}}_{\ell}(t) &= \gamma_1 \sum_{i=1}^{N} \left| (\mathrm{D}\bar{c})_i + t\,(\mathrm{D}\bar{z}_c)_i \right| + \left\| \bar{g} + t\,\bar{z}_g \right\|_{\infty} \\
&\quad + \frac{\gamma_2}{2} \left\| \left(\mathrm{Q}\bar{x}^{\perp} - \mathrm{H}f \right) + t \left(\mathrm{Q}\,\bar{z}^{\perp} \right) \right\|_2^2 \\
&= \gamma_1 \sum_{i=1}^{N} \left| \,(\mathrm{D}\bar{c})_i + t\,(\mathrm{D}\bar{z}_c)_i \,\right| + \max_{i=1,\ldots,N} \left| \,\bar{g}_i + t\,\bar{z}_{g,i} \,\right| \\
&\quad + \frac{\gamma_2}{2} \left(t^2 \left\| \mathrm{Q}\,\bar{z}^{\perp} \right\|_2^2 + 2t \left(\mathrm{Q}\bar{x}^{\perp} - \mathrm{H}f \right)^{\mathrm{T}} \mathrm{Q}\,\bar{z}^{\perp} + \left\| \mathrm{Q}\bar{x}^{\perp} - \mathrm{H}f \right\|_2^2 \right).
\end{aligned} \tag{93}$$

Since by definition of $\bar{z}^{\perp}$ we have that $\bar{z}^{\perp} \neq 0_N \Longrightarrow \mathrm{Q}\bar{z}^{\perp} \neq 0_N \Longrightarrow \|\mathrm{Q}\bar{z}^{\perp}\|_2 \neq 0$, then if the direction $\bar{z}$ of the line $x(t)$ has a non-null projection $\bar{z}^{\perp}$ onto the set $\mathcal{N}^{\perp}$ in (91), then the associated restriction $\widehat{\mathcal{J}}_{\ell}(t)$ in (93) is strongly convex in t and, hence, admits a unique global minimizer. This means that the restrictions of $\widetilde{\mathcal{J}}$ to all lines $x(t) \subset \mathcal{H}_v$ not parallel to the set $\mathcal{N}$ in (19) can intersect the set $\mathcal{M}$ of global minimizers of $\widetilde{\mathcal{J}}$ at most in one point. It follows that, if the set $\mathcal{M}$ is not a singleton, then it must be parallel to $\mathcal{N}$ and, hence, its maximum dimension is the dimension r of $\mathcal{N}$ given in (21). □

Acknowledgements This work was supported in part by the National Group for Scientific Computation (GNCS-INDAM), INdAM-GNCS Project 2023 code CUP_E53C22001930001, and in part by MIUR RFO projects. This study was carried out within the MICS (Made in Italy - Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) - MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 - D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.
This study was carried out within the MICS (Made in Italy - Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PNRR) - D.D. 1551.11-10-2022, PE00000004). This work was supported in part by the National Group for Scientific Computation (GNCS-INDAM), Research Projects 2024, and in part by MIUR RFO projects. The work of SM, MH and LG

was supported by PRIN2022_MORIGI, titled “Inverse Problems in the Imaging Sciences (IPIS)” 2022 ANC8HL - CUP J53D23003670006, and PRIN2022_PNRR_CRESCENTINI CUP.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Data Availibility Data sets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or nonfinancial interests to disclose.

Ethical Approval Not applicable.

References

1. Aujol, J.-F., Aubert, G., Blanc-Féraud, L., Chambolle, A.: Image decomposition into a bounded variation component and an oscillating component. J. Math. Imaging Vis. **22**, 71–88 (2005)
2. Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press, Cambridge (1987)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends in Machine Learning. Now Publishers (2011)
4. Butcher, J.C.: Numerical Methods for Ordinary Differential Equations. Wiley, Hoboken (2016)
5. Cai, G., Selesnick, I.W., Wang, S., Dai, W., Zhu, Z.: Sparsity-enhanced signal decomposition via generalized minimax-concave penalty for gearbox fault diagnosis. J. Sound Vib. **432**, 213–234 (2018)
6. Chan, R., Lanza, A., Morigi, S., Sgallari, F.: Convex non-convex image segmentation. Numer. Math. **138**, 635–680 (2018)
7. Chen, P.-Y., Selesnick, I.W.: Group-sparse signal denoising: non-convex regularization, convex optimization. IEEE Trans. Signal Process. **62**(13), 3464–3478 (2014)
8. Cicone, A., Huska, M., Kang, S.-H., Morigi, S.: JOT: a variational signal decomposition into jump, oscillation and trend. IEEE Trans. Signal Process. **70**, 772–784 (2022)
9. Cicone, A., Huska, M., Kang, S.-H., Morigi, S.: A two-stage signal decomposition into jump, oscillation and trend using ADMM. Image Process. On Line **13**, 153–166 (2023)
10. Cicone, A., Liu, J., Zhou, H.: Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis. Appl. Comput. Harmon. Anal. **41**(2), 384–411 (2016). Sparse Representations with Applications in Imaging Science, Data Analysis, and Beyond, Part II.
11. Condat, L.: A direct algorithm for 1-D total variation denoising. IEEE Signal Process. Lett. **20**(11), 1054–1057 (2013)
12. Condat, L.: Fast projection onto the simplex and the l_1 ball. Math. Program. **158**(1), 575–585 (2016)
13. Daubechies, I., Lu, J., Wu, H.-T.: Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. Appl. Comput. Harmon. Anal. **30**(2), 243–261 (2011)
14. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. Ser. B (Methodol.) **39**(1), 1–22 (1977)
15. Dragomiretskiy, K., Zosso, D.: Variational mode decomposition. IEEE Trans. Signal Process. **62**(3), 531–544 (2014)

16. Girometti, L., Huska, M., Lanza, A., Morigi, S.: Quaternary image decomposition with cross-correlation-based multi-parameter selection. In: Calatroni, L., Donatelli, M., Morigi, S., Prato, M., Santacesaria, M., (eds.) Scale Space and Variational Methods in Computer Vision: 9th International Conference, SSVM 2023, Santa Margherita Di Pula, Italy, May 21–25, 2023, Proceedings, pp. 120–133. Springer, Berlin (2023)
17. Girometti, L., Lanza, A., Morigi, S.: Ternary image decomposition with automatic parameter selection via auto- and cross-correlation. Adv. Comput. Math. **49**(1), 1 (2023)
18. Hiriart-Urruty, J., Lemarechal, C.: Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods. Grundlehren der mathematischen Wissenschaften. Springer, Berlin (1996)
19. Huang, G., Lanza, A., Morigi, S., Reichel, L., Sgallari, F.: Majorization–minimization generalized krylov subspace methods for ℓ_p-ℓ_q optimization applied to image restoration. BIT Numer. Math. **57**(2), 351–378 (2017)
20. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **454**(1971), 903–995 (1998)
21. Jung, M., Kang, M.: Efficient nonsmooth nonconvex optimization for image restoration and segmentation. J. Sci. Comput. **62**, 336–370 (2015)
22. Lanza, A., Pragliola, M., Sgallari, F.: Parameter-free restoration of piecewise smooth images. ETNA **59**, 202–229 (2023)
23. Lanza, A., Morigi, S., Sgallari, F.: Automatic parameter selection based on residual whiteness for convex non-convex variational restoration. In: Tai, X.-C., Wei, S., Liu, H. (eds.) Mathematical Methods in Image Processing and Inverse Problems, vol. 360. Springer, Singapore (2021)
24. Lanza, A., Morigi, S., Sgallari, F.: Convex image denoising via non-convex regularization with parameter selection. J Math Imaging Vis **56**, 195–220 (2016)
25. Lanza, A., Morigi, S., Selesnick, I.W., Sgallari, F.: Convex non-convex variational models. In: Chen, K., Schönlieb, C.-B., Tai, X.-C., Younces, L. (eds.) Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, pp 1–57. Springer, Cham (2021)
26. Lin, L., Wang, Y., Zhou, H.: Iterative filtering as an alternative algorithm for empirical mode decomposition. Adv. Adapt. Data Anal. **1**, 543–560 (2009)
27. Lorenz, D., Tran-Dinh, Q.: Non-stationary Douglas–Rachford and alternating direction method of multipliers: adaptive step-sizes and convergence. Comput. Optim. Appl. **74**, 67–92 (2019)
28. Lou, Y., Yin, P., Xin, J.: Point source super-resolution via non-convex based methods. J. Sci. Comput. **68**, 1082–1100 (2016)
29. Meyer, Y., Lewis, D.: Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures. Memoirs of the American Mathematical Society, American Mathematical Society (2001)
30. Meyers, B.E., Boyd, S.P.: Signal Decomposition Using Masked Proximal Operators. Foundations and Trends in Signal Processing. Now Publishers, Norwell (2023)
31. Nikolova, M., Ng, M.K., Tam, C.-P.: Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. IEEE Trans. Image Process. **19**(12), 3073–3088 (2010)
32. Liao, S., Fu, S., Li, Y., Han, H.: Image inpainting using non-convex low rank decomposition and multidirectional search. Appl. Math. Comput. **452**, 128048 (2023)
33. Selesnick, I., Lanza, A., Morigi, S., Sgallari, F.: Non-convex total variation regularization for convex denoising of signals. J. Math. Imaging Vis. **62**(6–7), 825–841 (2020)
34. Tang, L., Wu, L., Fang, Z., Li, C.: A non-convex ternary variational decomposition and its application for image denoising. IET Signal Process. **16**(3), 248–266 (2022)
35. Themelis, A., Patrinos, P.: Douglas–Rachford splitting and ADMM for nonconvex optimization: tight convergence results. SIAM J. Optim. **30**(1), 149–181 (2020)
36. Thi, H.A.L., Dinh, T.P.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**, 23–46 (2005)
37. Wu, T., Shao, J., Gu, X., Ng, M.K., Zeng, T.: Two-stage image segmentation based on nonconvex $\ell_2 - \ell_p$ approximation and thresholding. Appl. Math. Comput. **403**, 126168 (2021)
38. Yong, X., Ward, R.K., Birch, G.E.: Generalized Morphological Component Analysis for EEG source separation and artifact removal. In: 2009 4th International IEEE/EMBS Conference on Neural Engineering, pp. 343–346 (2009)

39. Yuille, A.L., Rangarajan, A.: The Concave-Convex Procedure (CCCP). In: Dietterich, T., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems, vol. 14. MIT Press, Cambridge (2001)