Computational approaches to non-convex, sparsity-inducing multi-penalty regularization

OPEN ACCESS
IOP Publishing
Inverse Problems
Inverse Problems **37** (2021) 055008 (22pp)
https://doi.org/10.1088/1361-6420/abdd46

Computational approaches to non-convex, sparsity-inducing multi-penalty regularization

Željko Kereta[1,2,*], Johannes Maly[3] and Valeriya Naumova[2]

[1] University College London, United Kingdom
[2] Simula Research Laboratory, Simula Metropolitan Center for Digital Engineering, Norway
[3] RWTH Aachen University, Germany

E-mail: zeljko@simula.no, maly@mathc.rwth-aachen.de and valeriya@simula.no

Received 23 September 2020, revised 10 January 2021
Accepted for publication 19 January 2021
Published 20 April 2021

Abstract

In this work we consider numerical efficiency and convergence rates for solvers of non-convex multi-penalty formulations when reconstructing sparse signals from noisy linear measurements. We extend an existing approach, based on reduction to an augmented single-penalty formulation, to the non-convex setting and discuss its computational intractability in large-scale applications. To circumvent this limitation, we propose an alternative single-penalty reduction based on infimal convolution that shares the benefits of the augmented approach but is computationally less dependent on the problem size. We provide linear convergence rates for both approaches, and their dependence on design parameters. Numerical experiments substantiate our theoretical findings.

Keywords: multi-penalty regularization, iterative thresholding, non-convex optimization, l_q-regularization ($0 < q < 1$)

(Some figures may appear in colour only in the online journal)

*Author to whom any correspondence should be addressed.

1. Introduction

In many real-life applications one is interested in recovering a structured signal from few corrupted linear measurements. One particular challenge lies in separating the ground-truth from pre-measurement noise since any such corruption is amplified during the measurement process, a phenomenon known as noise folding [2] or input noise model [1]. It commonly appears in signal processing and compressed sensing applications, where noise is added to the signal both before and after the measurement process occurs. This can be modeled as

$$\mathbf{A}(\mathbf{u}^\dagger + \mathbf{v}) + \boldsymbol{\xi} = \mathbf{y}, \tag{1}$$

where $\mathbf{u}^\dagger \in \mathbb{R}^n$ is an s-sparse original signal that we want to recover, $\mathbf{v} \in \mathbb{R}^n$ is the pre-measurement noise, $\boldsymbol{\xi} \in \mathbb{R}^m$ is the post-measurement noise, and $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the measurement matrix. Note that a signal $\mathbf{u} \in \mathbb{R}^n$ is called s-sparse if its support consists of at most s elements, i.e. $|\mathrm{supp}(\mathbf{u})| = |\{i : \mathsf{u}_i \neq 0\}| \leqslant s$. Information theoretic bounds state that the number of measurements m required for the exact support recovery of $\mathbf{u}^\dagger$ from (1) needs to scale linearly[4] with n, which leads to poor compression performance [1].

A number of recent studies [3, 15, 16, 21] try and mitigate these issues through a multi-penalty regularization framework defined as

$$\min_{\mathbf{u},\mathbf{v}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}(\mathbf{u}+\mathbf{v}) - \mathbf{y}\|_2^2 + \frac{\alpha}{q}\|\mathbf{u}\|_q^q + \frac{\beta}{p}\|\mathbf{v}\|_p^p, \tag{2}$$

where $\alpha, \beta > 0$ are regularization parameters, $0 \leqslant q < 2$, and $2 \leqslant p < \infty$. In particular, to promote sparsity of the $\mathbf{u}$ component we choose $q \leqslant 1$. A natural way to minimize (2) is via alternating minimization, starting from $\mathbf{u}^0, \mathbf{v}^0 \in \mathbb{R}^n$ and then iterating as

$$\begin{aligned} \mathbf{u}^{k+1} &\in \operatorname*{arg\,min}_{\mathbf{u}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}(\mathbf{u}+\mathbf{v}^k) - \mathbf{y}\|_2^2 + \frac{\alpha}{q}\|\mathbf{u}\|_q^q, \\ \mathbf{v}^{k+1} &\in \operatorname*{arg\,min}_{\mathbf{v}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}(\mathbf{u}^{k+1}+\mathbf{v}) - \mathbf{y}\|_2^2 + \frac{\beta}{p}\|\mathbf{v}\|_p^p. \end{aligned} \tag{3}$$

Whereas the second problem is differentiable and admits an explicit solution, the first problem requires iterative thresholding for $q \leqslant 1$ [21], for each outer iteration $k \in \mathbb{N}$, and becomes non-convex if $q < 1$. Moreover, alternating minimization does not lend itself to an easy analysis of the convergence rate.

1.1. Contribution

In this work we examine the multi-penalty problem (2), for the case $0 < q \leqslant 1$ and $p = 2$. We first show that the augmented approach in [16], which allows to decouple the computation of $\mathbf{u}$ and $\mathbf{v}$ components of the solution, can be easily extended to $q < 1$ to obtain an augmented single-penalty iterative thresholding algorithm providing solutions to (2). Since this includes computing the inverse of a possibly high-dimensional matrix, we suggest an alternative single-penalty iterative thresholding algorithm which is based on an infimal convolution formulation

[4] Assume for simplicity $\mathbf{v} \perp\!\!\!\perp \boldsymbol{\xi}$, $\boldsymbol{\xi} \sim \mathcal{N}(0, \sigma^2\,\mathbf{Id}_n)$, and $\mathbf{v} \sim \mathcal{N}(0, \sigma_v^2\,\mathbf{Id}_n)$. We now write (1) as $\mathbf{y} = \mathbf{A}\mathbf{u}^\dagger + \mathbf{w}$, where $\mathbf{w} := \mathbf{A}\mathbf{v} + \boldsymbol{\xi}$ represents the effective noise. The covariance matrix of $\mathbf{w}$ equals $\sigma^2\,\mathbf{Id}_m + \sigma_v^2\mathbf{A}\mathbf{A}^\top =: \mathbf{Q}$. Assuming $\mathbf{A}\mathbf{A}^\top \approx \frac{n}{m}\,\mathbf{Id}_m$ (as is the case, with high probability, for $\mathbf{A}$ with zero mean, $1/m$-variance sub-Gaussian entries), and $\sigma_v \approx \sigma$, we would have $\mathbf{Q} = \sigma^2(1 + C\frac{n}{m})\mathbf{Id}_m$, for $C > 0$. Thus, the variance of the noise rises by a factor proportional to n/m, which when $m \ll n$ can be substantial.

of (2) and sidesteps the computational bottleneck of the augmented approach. We show a linear convergence rate for both approaches, in dependence of design parameters, and in numerical simulations confirm both the rate analysis and the efficiency gap. In particular, we argue that the benefits of faster convergence rates are sometimes offset by the computational demands, which suggests that a preferred method for solving the optimization problem can be chosen with respect to the size of $\mathbf{A}$.

1.2. Related work

In [21] the authors approach (3), for $0 < q \leqslant 1$ and $p = \infty$, on separable Hilbert spaces by applying iterative thresholding algorithms to each of the sub-problems, and show convergence of the sequence of iterates to stationary points of the underlying problem. The choice $p = \infty$ is of special interest when $\mathbf{v}$ models uniform pre-measurement noise. However, the authors also show that $p = 2$ exhibits the best (empirical) performance for the reconstruction of $\mathbf{u}^\dagger$, for $\mathbf{v}$ modelling various common noise types (including uniform noise). It is for this reason that in this paper we are concerned only with the case $p = 2$. We add though that more general noise types might be of interest in very particular cases, and this is a possible topic for future research. In [16] the authors reduce the optimization problem (2) to a single-penalty regularization through an augmented data matrix, for $q = 1$ and $p = 2$, and derive conditions on optimal support recovery. The authors provide theoretical and numerical evidence of superior performance of multi-penalty regularization over standard single-penalty approaches for the sparse recovery of solutions to (1). In [15] a principled, data-driven parameter selection approach is derived for $q = 1$ and $p = 2$, based on the Lasso path. Instead of through noise folding, a multi-penalty formulation of the objective function can also be seen from the perspective of the recovery of a signal that is a superposition of two components, e.g. a sparse and a smooth component. See [12] and references therein. In spite of these and other advances, rigorous results regarding convergence rate and error analysis for (2) have not been established.

Since we reduce (2) to specific single-penalty problems, corresponding convergence results on classical proximal descent methods are of interest. In [9] important insights on support stability and convergence of iterative thresholding algorithms on separable Hilbert spaces have been collected while [28] proved linear convergence rates of the iterative thresholding algorithm, under certain conditions, if the underlying thresholding operator is not continuous, though the dependency on the parameters of the optimization scheme are not explicitly derived. Linear convergence of a single penalty non-convex regularizer with adaptive thresholding was established in [24], where the influence of the restricted isometry property (RIP) of the design matrix on the convergence constant can be inferred. A further survey of nonconvex regularizers for sparse recovery can be found in [25].

Lastly, approaches representing regularizers as infimal convolution can be found in the context of machine learning and signal processing, cf [17, 18]. Therein primal-dual schemes are examined for optimizing functionals penalized via infimal convolutions. The results, however, require piece-wise convexity which is not given in our case.

1.3. Notation

We restrict boldface lettering to matrices (uppercase), e.g. $\mathbf{A}$, and vectors (lowercase), e.g. $\mathbf{u}$. The ith entry of a vector $\mathbf{u}$ is denoted as u_i. For $m \in \mathbb{N}$ we denote $[m] := \{1, \ldots, m\}$. For $0 < q \leqslant \infty$ the ℓ_q norm of a vector $\mathbf{u} = (\mathsf{u}_1, \ldots, \mathsf{u}_n)^\top \in \mathbb{R}^n$ is denoted by $\|\mathbf{u}\|_q$. The support

set of $\mathbf{u} \in \mathbb{R}^n$ is denoted as

$$\operatorname{supp}(\mathbf{u}) = \{i \in [n] : \mathsf{u}_i \neq 0\}$$

and the sign $\operatorname{sgn}(\mathbf{u}) = (\operatorname{sgn}(\mathsf{u}_i))_{i=1}^n$ is defined component-wise by

$$\operatorname{sgn}(\mathsf{u}) = \begin{cases} 1, & \text{if } \mathsf{u} > 0, \\ 0, & \text{if } \mathsf{u} = 0, \\ -1, & \text{if } \mathsf{u} < 0. \end{cases}$$

For a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, we use $\|\mathbf{M}\|$ to denote its spectral norm and $\lambda_{\min}(\mathbf{M})$ to denote its smallest singular value. We denote the $n \times n$ identity matrix by $\mathbf{Id}_n$. For $I \subset [n]$, $\mathbf{M}_I \in \mathbb{R}^{m\times|I|}$ represents the submatrix of $\mathbf{M}$ containing the columns indexed by I, and $\mathbf{u}_I \in \mathbb{R}^{|I|}$ denotes the subvector of $\mathbf{u}$ containing the entries restricted to I. We denote the corresponding orthogonal projection operator onto I as $\mathbf{P}_I \in \mathbb{R}^{|I|\times n}$, so that $\mathbf{P}_I\mathbf{u} = \mathbf{u}_I$. When indexed by a set $T \subset \mathbb{R}^n$, $\mathbf{P}_T$ denotes the orthogonal projection onto T. Finally, the set-valued operator ∂ denotes the limiting Fréchet subdifferential, and $\operatorname{dom}\partial f = \{\mathbf{x} : \partial f(\mathbf{x}) \neq \emptyset\}$ is its corresponding domain when applied to a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, cf [20, 23].

2. Main results

Consider the multi-penalty problem (2) for $p = 2$, i.e. minimizing

$$\mathcal{T}^q_{\alpha,\beta}(\mathbf{u}, \mathbf{v}) := \frac{1}{2}\|\mathbf{A}(\mathbf{u} + \mathbf{v}) - \mathbf{y}\|_2^2 + \frac{\alpha}{q}\|\mathbf{u}\|_q^q + \frac{\beta}{2}\|\mathbf{v}\|_2^2, \tag{4}$$

and denote a corresponding solution pair by

$$\left(\mathbf{u}^q_{\alpha,\beta}, \mathbf{v}^q_{\alpha,\beta}\right) \in \operatorname*{arg\,min}_{\mathbf{u},\mathbf{v}\in\mathbb{R}^n} \mathcal{T}^q_{\alpha,\beta}(\mathbf{u}, \mathbf{v}). \tag{5}$$

As mentioned above $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{y} \in \mathbb{R}^m$, $\alpha, \beta > 0$ are regularization parameters balancing the contributions of the data-fidelity term and the two regularization terms, and $0 < q \leqslant 1$.

Let us introduce two widely known concepts relevant for the forthcoming discussion. First, the *Kurdyka–Łojasiewicz* (KŁ) property; a well-established tool for analyzing the convergence, and convergence rates, of proximal descent algorithms [4].

Definition 2.1. A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is said to have the KŁ property at $\bar{\mathbf{x}} \in \operatorname{dom}\partial f$ if there exists $\eta \in (0, +\infty]$, a neighbourhood Ω of $\bar{\mathbf{x}}$, and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$ such that

(a) $\varphi \in C^1(0, \eta)$, $\varphi(0) = 0$ and $\varphi'(s) > 0$ for all $s \in (0, \eta)$
(b) For all $\mathbf{x} \in \Omega \cap \{\mathbf{x} : f(\bar{\mathbf{x}}) < f(\mathbf{x}) < f(\bar{\mathbf{x}}) + \eta\}$ the KŁ inequality holds

$$\varphi'(f(\mathbf{x}) - f(\bar{\mathbf{x}}))\operatorname{dist}(0, \partial f(\mathbf{x})) \geqslant 1.$$

The KŁ property is used to describe the speed of convergence through the desingularizing function φ. It has been shown that semi-algebraic functions satisfy the KŁ property with $\varphi(s) = cs^{1-\theta}$, where $c > 0$ and $\theta \in [0, 1)$ is called the KŁ constant, which characterizes the convergence speed of proximal gradient descent algorithms [4, theorem 11]. As observed in [8], corollary 3.6 in [19] may be used to determine the KŁ constant of piecewise convex polynomials. Even though $\|\cdot\|_q^q$ has the KŁ property, cf [5, example 5.4], it does not result in

piece-wise convex polynomials for $0 < q < 1$, and thus we cannot apply [19, corollary 3.6] to infer the speed of convergence. We will instead adopt and adapt the ideas from [9, 28].

The second concept relevant for this paper is the RIP, which allows to control eigenvalues of small submatrices of $\mathbf{A} \in \mathbb{R}^{m\times n}$, and to characterize measurement operators that allow stable and robust reconstruction of sparse signals from $m \ll n$ measurements.

Definition 2.2. A matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ satisfies the restricted isometry property of order s (s-RIP) with constant $\delta_s \in (0, 1)$, if for all s-sparse $\mathbf{u} \in \mathbb{R}^n$

$$(1 - \delta_s)\|\mathbf{u}\|_2 \leqslant \|\mathbf{A}\mathbf{u}\|_2 \leqslant (1 + \delta_s)\|\mathbf{u}\|_2.$$

Remark 2.3. For a detailed treatment of RIP, and measurement operators that fulfill it, we refer the reader to [14]. Let us only mention that if the entries of $\mathbf{A}$ are i.i.d. copies of a Gaussian random variable with mean zero and variance $\frac{1}{m}$, then

$$m \geqslant C\delta_s^{-2} s \log\left(\frac{en}{s}\right)$$

measurements suffice to have an s-RIP with constant $\delta_s > 0$ with high probability, for an absolute constant $C > 0$. Consequently, $\delta_s = \mathcal{O}\left(m^{-1/2}\sqrt{s\,\log(en/s)}\right)$ with high probability.

2.1. Augmented formulation

It was observed in [16] that for $q = 1$, the multi-penalty problem (2) reduces to single-penalty regularization where measurement matrix and datum are adjusted by the regularization parameter β. We include this result, extended to $0 < q \leqslant 1$, together with the proof (see section A.1), which is analogous to [16, lemma 1].

Lemma 2.4. *The pair* $(\mathbf{u}^q_{\alpha,\beta}, \mathbf{v}^q_{\alpha,\beta})$ *minimizes* $\mathcal{T}^q_{\alpha,\beta}$ *in* (4) *if and only if*

$$\mathbf{v}^q_{\alpha,\beta} = v(\mathbf{u}^q_{\alpha,\beta}) = \left(\beta\,\mathbf{Id}_n + \mathbf{A}^\top\mathbf{A}\right)^{-1}\left(\mathbf{A}^\top\mathbf{y} - \mathbf{A}^\top\mathbf{A}\mathbf{u}^q_{\alpha,\beta}\right), \tag{6}$$

and $\mathbf{u}^q_{\alpha,\beta}$ *is the solution of the* augmented problem

$$\mathbf{u}^q_{\alpha,\beta} \in \operatorname*{arg\,min}_{\mathbf{u}\in\mathbb{R}^n} \mathcal{F}_\beta(\mathbf{u}), \quad \mathcal{F}_\beta(\mathbf{u}) := \frac{1}{2}\left\|\mathbf{B}_\beta\mathbf{u} - \mathbf{y}_\beta\right\|_2^2 + \frac{\alpha}{q}\|\mathbf{u}\|_q^q, \tag{7}$$

with

$$\mathbf{B}_\beta = \left(\mathbf{Id}_m + \frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1/2}\mathbf{A} \quad \textit{and} \quad \mathbf{y}_\beta = \left(\mathbf{Id}_m + \frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1/2}\mathbf{y}.$$

Remark 2.5. The noise folding forward model (1) is in [2] written in the whitened form as $\tilde{\mathbf{y}} = \mathbf{B}\mathbf{u}^\dagger + \boldsymbol{\eta}$, for $\tilde{\mathbf{y}} = \mathbf{Q}^{-1/2}\mathbf{y}$, $\mathbf{B} = \mathbf{Q}^{-1/2}\mathbf{A}$, $\boldsymbol{\eta} = \mathbf{Q}^{-1/2}(\mathbf{A}\mathbf{v} + \boldsymbol{\xi})$, for $\mathbf{Q} = \frac{1}{c}(\sigma^2\,\mathbf{Id}_m + \sigma_v^2\mathbf{A}\mathbf{A}^\top)$ and $c > 0$ is a constant. Notice that this is particularly related to the augmented problem in (7). On an unrelated note, improving on the analysis in [2, proposition 2] one can show (see lemma B.1) that the coherence, defined for a matrix $\mathbf{M}$ as

$$\operatorname{coh}(\mathbf{M}) = \max_{i\neq j}\frac{\left|\mathbf{m}_i^\top\mathbf{m}_j\right|}{\|\mathbf{m}_i\|_2\|\mathbf{m}_j\|_2},$$

where $\mathbf{m}_i$ is the ith column of $\mathbf{M}$, of the augmented measurement matrix $\mathbf{B}_\beta$ satisfies

$$\operatorname{coh}(\mathbf{B}_\beta) \leqslant \left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)\left(\operatorname{coh}(\mathbf{A}) + \frac{\|\mathbf{A}\|^2}{\beta}\right). \tag{8}$$

In compressed sensing literature the magnitude of the coherence of a matrix is an important measure of quality for measurement matrices, cf [14, section 5]. The bound in (8) thus suggests that for small $\|\mathbf{A}\|$ or large β, the linear measurement process modelled by $\mathbf{B}_\beta$ is as information preserving as the one modelled by $\mathbf{A}$. In addition, lemma B.2 shows that $\operatorname{coh}(\mathbf{B}_\beta)$ behaves like the coherence of a conditioned version of $\mathbf{A}$ if $\beta \to 0$. Let us mention that in practice $\operatorname{coh}(\mathbf{B}_\beta)$ behaves well for all β's, and even moderate values of $\|\mathbf{A}\mathbf{A}^\top\|$.

By lemma 2.4, to estimate the solution pair $(\mathbf{u}^q_{\alpha,\beta}, \mathbf{v}^q_{\alpha,\beta})$ it is sufficient to first solve (7), and then insert the computed solution into (6). Since the fidelity term $\frac{1}{2}\|\mathbf{B}_\beta\mathbf{u} - \mathbf{y}_\beta\|_2^2$ is smooth and the regularization term $\|\mathbf{u}\|_q^q$ non-convex, the common approach is to use iterative thresholding through a forward-backward splitting algorithm [4, 9]. For $\mathcal{F}_\beta$ and the augmented problem (7), the resulting thresholding iterations applied are readily written as

$$\begin{cases} \text{Set the initial vector } \mathbf{u}^0 \\ \mathbf{u}^{k+1} = \operatorname{prox}_{\mu,\frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{u}^k - \mu\mathbf{B}_\beta^\top(\mathbf{B}_\beta\mathbf{u}^k - \mathbf{y}_\beta)). \end{cases} \tag{9}$$

Each iteration in (9) can be viewed as a thresholded Landweber iteration; we first perform a step in the direction of the negative gradient of the data fidelity term, and then apply the proximal operator of the remaining non-convex term.

The proximal operator of a function $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$\operatorname{prox}_{\mu,\nu\Psi}(\mathbf{u}) = \operatorname*{arg\,min}_{\mathbf{z}\in\mathbb{R}^n} \frac{1}{2\mu}\|\mathbf{z} - \mathbf{u}\|_2^2 + \nu\Psi(\mathbf{z}), \tag{10}$$

where $\mu, \nu > 0$. For separable mappings (10) can be applied component-wise, and we have $\operatorname{prox}_{\mu,\nu\|\cdot\|_q^q}(\mathbf{u}) = \left(\operatorname{prox}_{\mu,\nu|\cdot|^q}(\mathsf{u}_i)\right)_{i=1}^n$. In the general case, the proximal operator (10) could be set-valued, since there might be multiple or even no minima. It can be shown though that for $0 < q < 1$ the (one-dimensional) proximal operator of $|\cdot|^q$ satisfies

$$\operatorname{prox}_{\mu,\nu|\cdot|^q}(\mathsf{u}) = \begin{cases} \left(\cdot + \nu\mu q\ \operatorname{sgn}(\cdot)|\cdot|^{q-1}\right)^{-1}(\mathsf{u}), & \text{for } |\mathsf{u}| > \tau_\mu \\ 0, & \text{for } |\mathsf{u}| \leqslant \tau_\mu \end{cases}, \tag{11}$$

$$\text{where } \tau_\mu = \frac{2-q}{2-2q}(2\nu\mu(1-q))^{\frac{1}{2-q}}.$$

The range of $\operatorname{prox}_{\mu,\nu|\cdot|^q}$ is $(-\infty, -\lambda_{\mu,q}] \cup \{0\} \cup [\lambda_{\mu,q}, \infty)$ where $\lambda_{\mu,q} = (2\ \mu\nu(1-q))^{\frac{1}{2-q}}$, see [9, lemma 5.1], and it is discontinuous with a jump discontinuity[5] at $|\mathsf{u}| = \tau_\mu$. Note that the proximal operators in (11) are indeed thresholding operators, and as q goes from 0 to 1 they interpolate between hard- and soft-thresholding operators. Moreover, a closed form of the operator $\operatorname{prox}_{\mu,\nu|\cdot|^q}$ is known only in special cases, namely for $q = 1/2$ and $q = 2/3$ [26].

[5] While the actual proximal operator of $|\cdot|^q$ is set-valued and simultaneously assumes both possible values at $|\mathsf{u}| = \tau_\mu$, we follow common practice when restricting the operator to zero at $|\mathsf{u}| = \tau_\mu$ to have a single-valued function.

It follows easily that if the step-size $\mu > 0$ is small enough (smaller than $\|\mathbf{B}_\beta\|^{-2}$), the difference of iterates in (9) decreases, i.e. $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \to 0$ as $k \to \infty$, see [9, proposition 2.1]. Note that the iterations in (9) are quite different from those given by alternating minimization, where for each k we need to compute $\mathbf{u}^{k+1}$ through iterative thresholding. The following lemma makes this more precise; it shows that (9) is equivalent to performing only the first step of iterative thresholding when computing $\mathbf{u}^{k+1}$ in (3). The proof can be found in section A.2.

Lemma 2.6. *The iterations defined in* (9) *can be rewritten as*

$$\mathbf{u}^{k+1} = \operatorname{prox}_{\mu, \frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{u}^k - \mu \mathbf{A}^\top(\mathbf{A}\mathbf{u}^k + \mathbf{A}v(\mathbf{u}^k) - \mathbf{y})),$$

which corresponds to a single proximal gradient descent step of (3) *starting at* u^k.

2.1.1. Linear convergence. We now show that the iterates in (9) converge at a linear rate to stationary points $\mathbf{u}^\star$ of $\mathcal{T}^q_{\alpha,\beta}$, i.e. points such that $\mathbf{0} \in \partial \mathcal{T}^q_{\alpha,\beta}(\mathbf{u}^\star)$, and characterize the convergence constant in dependence of design parameters. Let us emphasize that since our analysis is tailored to ℓ_q-regularization we derive more explicit guarantees (in terms of the involved parameters) than what would follow by directly applying the more general statements of [28] to the augmented formulation (7). The proof can be found in section A.3.

Theorem 2.7. *Let* $\alpha, \beta > 0$ *and* $0 < q \leqslant 1$. *Assume the matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *has RIP of order* s *with a constant* $\delta_s \in (0,1)$, *and let the stepsize* μ *satisfy* $0 < \mu < \|\mathbf{A}\|^{-2} + \beta^{-1}$. *Moreover, assume*[6] $\mathbf{u}^\star \in \mathbb{R}^n$ *is such that* $|\mathrm{supp}(\mathbf{u}^\star)| \leqslant s$ *and the iterates* (9) *satisfy* $\mathbf{u}^k \to \mathbf{u}^\star$. *Define* $I = \mathrm{supp}(\mathbf{u}^\star)$ *and* $d_{\min} = \min_{i\in I}|\mathrm{u}_i^\star|$. *Then there exists* $k_0 \in \mathbb{N}$ *such that for all* $k \geqslant k_0$ *we have*

$$\|\mathbf{u}^{k+1} - \mathbf{u}^\star\|_2 \leqslant \frac{1 - \mu\left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1}(1-\delta_s)^2}{1 - \mu\alpha(1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}}\|\mathbf{u}^k - \mathbf{u}^\star\|_2.$$

Remark 2.8.

(a) To have linear convergence in theorem 2.7, we have to choose an α such that

$$0 < \alpha < \alpha^\star = \left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1}\frac{(1-\delta_s)^2}{(1-q)}\left(\frac{d_{\min}}{2}\right)^{2-q}. \tag{12}$$

This resembles basic assumptions of the main result in [28]. One should thus interpret theorem 2.7 as an additional refinement, better capable of predicting numerical behavior.

(b) Theorem 2.7 suggests that the convergence constant depends on the sparsity of the signal and properties of $\mathbf{A}$. Namely, if the signal is sparser (and thus δ_s smaller) then the convergence constant decreases. Similarly, the constant decreases if we increase the number of measurements.

(c) Assuming $\alpha = c\alpha^\star$, for $c \in (0,1)$, it is straight-forward to check that the rate in theorem 2.7 becomes minimal by choosing $\mu \approx \|\mathbf{A}\|^{-2} + \beta^{-1}$. In this case the result transforms into

$$\|\mathbf{u}^{k+1} - \mathbf{u}^\star\|_2 \leqslant \frac{1 - \|\mathbf{A}\|^{-2}(1-\delta_s)^2}{1 - c\|\mathbf{A}\|^{-2}(1-\delta_s)^2}\|\mathbf{u}^k - \mathbf{u}^\star\|_2.$$

[6] The sequence $\mathbf{u}^k$ converges provably to a stationary point since $\mathcal{T}^q_{\alpha,\beta}$ is among other things coercive and has the KL-property, cf [5, theorem 5.1]. The assumption thus is not about whether $\mathbf{u}_k$ converges but about the specific limit point which mainly depends on the concrete choice of initialization.

(d) Since α and β control the strength of regularization in $\mathcal{T}^q_{\alpha,\beta}$, their choice depends on the expected noise level. Consequently, when setting α and β one needs to make a trade-off between their regularizing effect and the desired convergence speed.

2.1.2. Computational complexity. Once $\mathbf{B}_\beta$ has been computed, executing (9) for a constant number of iterations costs $\mathcal{O}(mn)$ operations: $\mathcal{O}(mn)$ for matrix-vector products and $\mathcal{O}(n)$ for evaluating the proximal operator. But this gets dominated by the operations needed to obtain $\mathbf{B}_\beta$, which involve a matrix square root and a matrix–matrix linear system and have to be done in advance. This turns out to be a computational bottleneck as soon as $m \geqslant n^{\frac{1}{\rho-1}}$ as it requires $\mathcal{O}(m^\rho)$ operations, where $\rho \in [2.37, 3]$ depends on the used algorithmic method [11]. Such a computational cost can be prohibitive for high-dimensional applications.

2.2. Infimal convolution formulation

To overcome the computational limitations observed above, we consider an alternative approach. Define a new program by

$$\mathbf{w}^q_{\alpha,\beta} = \operatorname*{arg\,min}_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}\mathbf{w}-\mathbf{y}\|_2^2 + \left(\frac{\alpha}{q}\|\cdot\|_q^q \Delta \frac{\beta}{2}\|\cdot\|_2^2\right)(\mathbf{w}), \tag{13}$$

where the infimal convolution is given by

$$g(\mathbf{w}) := \left(\frac{\alpha}{q}\|\cdot\|_q^q \Delta \frac{\beta}{2}\|\cdot\|_2^2\right)(\mathbf{w}) = \inf_{\mathbf{u}\in\mathbb{R}^n} \frac{\alpha}{q}\|\mathbf{u}\|_q^q + \frac{\beta}{2}\|\mathbf{w}-\mathbf{u}\|_2^2. \tag{14}$$

For a detailed treatment of infimal convolution and its properties, see [6]. It is straight-forward to check that an equivalence between minimizing (4) and (13) holds.

Lemma 2.9. *The pair* $(\mathbf{u}^q_{\alpha,\beta}, \mathbf{v}^q_{\alpha,\beta})$ *minimizes* $\mathcal{T}^q_{\alpha,\beta}$ *in* (4) *if and only if* $\mathbf{u}^q_{\alpha,\beta} + \mathbf{v}^q_{\alpha,\beta}$ *solves* (13) *while* $\mathbf{u}^q_{\alpha,\beta}$ *attains the infimal value of* $\left(\frac{\alpha}{p}\|\cdot\|_q^q \Delta \frac{\beta}{2}\|\cdot\|_2^2\right)(\mathbf{u}^q_{\alpha,\beta} + \mathbf{v}^q_{\alpha,\beta})$.

In order to solve (13) via iterative thresholding (i.e. proximal gradient descent), we need to efficiently evaluate the proximal operator of (14). A helpful observation is that (14) can be interpreted as the Moreau-envelope of $\|\cdot\|_q^q$, which for a function f and $t > 0$ is defined as

$$M_{t,f}(\mathbf{x}) = \left(f \Delta \frac{1}{2t}\|\cdot\|_2^2\right)(\mathbf{x}) = f(\operatorname{prox}_{t,f}(\mathbf{x})) + \frac{1}{2t}\|\mathbf{x} - \operatorname{prox}_{t,f}(\mathbf{x})\|_2^2,$$

where the last equality only holds if $\operatorname{prox}_{tf}(\mathbf{x}) \neq \emptyset$. It has been observed in [7, theorem 6.63] that computing the proximal operator of the Moreau envelope reduces to computing the proximal operator of the underlying function. Though stated only for convex functions in [7], it is straight-forward to generalize the result.

Lemma 2.10. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a lower semi-continuous function with* $f(0) = \min f$. *Then,*

$$\operatorname{prox}_{\mu,\lambda M_{t,f}}(\mathbf{x}) = \frac{t}{t+\mu\lambda}\mathbf{x} + \frac{\mu\lambda}{t+\mu\lambda}\operatorname{prox}_{(t+\mu\lambda),f}(\mathbf{x}).$$

The proof is in section A.4. Define now the proximal gradient descent for (13) by

$$\begin{cases} \text{Set the initial vector } \mathbf{w}^0 \\ \mathbf{w}^{k+1} = \operatorname{prox}_{\mu,g}(\mathbf{w}^k - \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y})). \end{cases} \tag{15}$$

We denote by $\mathbf{u}^k = \operatorname{prox}_{\frac{1}{\beta},\frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{w}^k)$ the sequence of minimizers attaining $g(\mathbf{w}^k)$, and set $\mathbf{v}^k = \mathbf{w}^k - \mathbf{u}^k$. Note that with this notation $\mathbf{w}^k$ and $\mathbf{u}^k$ can also be characterized via

$$\begin{cases} \mathbf{w}^k = \underset{\mathbf{w}\in\mathbb{R}^n}{\arg\min}\, \dfrac{1}{2\mu}\|\mathbf{w} - \mathbf{w}^{k-1} + \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^{k-1} - \mathbf{y})\|_2^2 + \dfrac{\beta}{2}\|\mathbf{w} - \mathbf{u}^k\|_2^2 \\ \mathbf{u}^k = \underset{\mathbf{u}\in\mathbb{R}^n}{\arg\min}\, \dfrac{\beta}{2}\|\mathbf{u} - \mathbf{w}^k\|_2^2 + \dfrac{\alpha}{q}\|\mathbf{u}\|_q^q. \end{cases} \tag{16}$$

Unlike (15), the representation in (16) does not yield a practically viable algorithm, since $\mathbf{w}^k$ and $\mathbf{u}^k$ are not decoupled. It does though lend itself to theoretical analysis of the iterations, cf section A.5.

2.2.1. Linear convergence. Though g in (14) is continuous and separable, i.e. $g(\mathbf{w}) = \sum_{i=1}^n g_i(\mathsf{w}_i)$, it is not continuously differentiable, such that we cannot apply [28] to deduce linear convergence of (15). Nevertheless, using the KKT-conditions of the objective functions in (16), we get linear convergence of the iterates in (15) by a similar strategy as in theorem 2.7.

Theorem 2.11. *Let $\alpha, \beta > 0$ and $0 < q \leqslant 1$. Assume*[7] *that $0 < \mu < \|\mathbf{A}\|^{-2}$ and $\mathbf{w}^k \to \mathbf{w}^\star$. Let $I \subset [n]$ denote the support of $\mathbf{u}^\star = \operatorname{prox}_{\frac{1}{\beta},\frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{w}^\star)$ and define $d_{\min} = \min_{i\in I}|\mathsf{u}_i^\star|$. Then there exists $k_0 \in \mathbb{N}$ such that for all $k \geqslant k_0$ we have*

$$\|\mathbf{w}^{k+1} - \mathbf{w}^\star\|_2 \leqslant \left(\frac{\|\mathbf{P}_I - \mu\mathbf{A}_I^\top\mathbf{A}\|^2}{\left(1 - \alpha\mu(1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}\right)^2} + \frac{\|\mathbf{P}_{I^c} - \mu\mathbf{A}_{I^c}^\top\mathbf{A}\|^2}{(1+\mu\beta)^2} \right)^{1/2} \|\mathbf{w}^k - \mathbf{w}^\star\|_2$$

The proof of theorem 2.11 is given in section A.5.

Remark 2.12. On the one hand, in theorem 2.11 the assumption on μ and the rate differ from theorem 2.7; there is no influence of β on admissible step-sizes and the rate is split in two distinct components. On the other hand, since, for $\mu < \|\mathbf{A}\|^{-2}$,

$$\begin{aligned} \|\mathbf{P}_I - \mu\mathbf{A}_I^\top\mathbf{A}\| = \|\mathbf{P}_I(\mathbf{Id}_n - \mu\mathbf{A}^\top\mathbf{A})\| \leqslant \|\mathbf{Id}_n - \mu\mathbf{A}^\top\mathbf{A}\| < 1 \text{ and} \\ \|\mathbf{P}_{I^c} - \mu\mathbf{A}_{I^c}^\top\mathbf{A}\| = \|\mathbf{P}_{I^c}(\mathbf{Id}_n - \mu\mathbf{A}^\top\mathbf{A})\| \leqslant \|\mathbf{Id}_n - \mu\mathbf{A}^\top\mathbf{A}\| < 1, \end{aligned} \tag{17}$$

the rate in theorem 2.11 suggests to choose β large to dominate the second term of the rate in which case the assumptions on μ agree in both theorems. Moreover, this reduces the rate to

$$\|\mathbf{w}^{k+1} - \mathbf{w}^\star\|_2 \leqslant \left(\frac{\|\mathbf{P}_I - \mu\mathbf{A}_I^\top\mathbf{A}\|}{1 - \alpha\mu(1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}} + \mathcal{O}(\beta^{-1}) \right) \|\mathbf{w}^k - \mathbf{w}^\star\|_2,$$

where the denominator is as in theorem 2.7. In light of (17), we get linear convergence of (15) if

$$0 < \alpha < \alpha^* = \frac{1 - \|\mathbf{P}_I - \mu\mathbf{A}_I^\top\mathbf{A}\|}{\mu(1-q)}\left(\frac{d_{\min}}{2}\right)^{2-q}.$$

As already discussed in remark 2.8, a trade-off between regularization and convergence rate has to be taken into account when choosing α and β.

[7] Along the lines of footnote 6 in theorem 2.7. Just note that g in (14) has the KL-property by [27, theorem 3.1] and, hence, the objective function in (13) has it as well.

Remark 2.13. For $q = 1$, an alternative viewpoint on (16) is given by

$$\begin{aligned}\mathbf{w}^{k+1} &= \operatorname*{arg\,min}_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2\mu}\|\mathbf{w} - \mathbf{w}^k + \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y})\|_2^2 + \frac{\beta}{2}\|\mathbf{w} - \mathbf{u}^{k+1}\|_2^2 \\ &= \operatorname*{arg\,min}_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2\mu}\|\mathbf{w} - \mathbf{w}^k + \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y})\|_2^2 + \frac{\beta}{2}\|\mathbf{w} - \operatorname{prox}_{\frac{\alpha}{\beta}\|\cdot\|_1}(\mathbf{w})\|_2^2 \\ &= \operatorname*{arg\,min}_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2\mu}\|\mathbf{w} - \mathbf{w}^k + \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y})\|_2^2 + \frac{\alpha}{2}\|\nabla M_{\frac{\alpha}{\beta}\|\cdot\|_1}(\mathbf{w})\|_2^2,\end{aligned} \tag{18}$$

where we used [22, equation (3.3)] in the last step, meaning that

$$\mathbf{w}^{k+1} = \operatorname{prox}_{\frac{\alpha\mu}{2}\|\nabla M_{\frac{\alpha}{\beta}\|\cdot\|_1}(\cdot)\|_2^2}(\mathbf{w}^k - \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y}))$$

is a proximal gradient descent sequence of $\|\nabla M_{\frac{\alpha}{\beta}\|\cdot\|_1}(\cdot)\|_2^2$, the squared ℓ_2-norm of the gradient of the smooth Moreau approximation of $\frac{\alpha}{\beta}\|\cdot\|_1$. From this perspective, multi-penalty regularization resembles a Newton-type method by searching for zeros of the derivative of a smooth approximation of the ℓ_1-norm. However, transferring this intuition to the case $q < 1$ is non-trivial. On a technical level the equations in (18) break down in the third line, which does not hold for $q < 1$ due to non-convexity of $\|\cdot\|_q^q$.

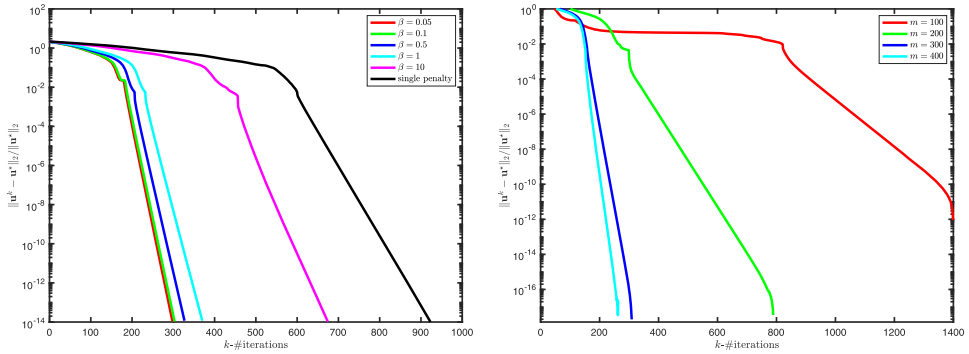
2.2.2. Computational complexity. While (9) requires computing $\mathbf{B}_\beta$, which can be costly, the infimal convolution formulation (15) does not incur additional computational costs and thus directly inherits efficiency and linear convergence of the proximal descent method. Indeed, for a fixed number of iterations the number of operations performed in (15) is $\mathcal{O}(mn)$ (the additional convex combination when evaluating the proximal operator by lemma 2.10 is negligible). This is considerably lower than $\mathcal{O}(m^\rho)$, for $\rho \in [2.37, 3]$, which is the computational cost of the augmented formulation, particularly if m is large. In numerical simulations, this effect is easy to observe, cf section 3.

3. Numerical experiments

We now present experimental results that focus on two aspects of our study. First, we examine the convergence rate of the proposed algorithms, confirming linear convergence and in case of the augmented formulation, the dependence of the convergence constant on the parameters of the problem. Second, we examine their efficiency by studying the overall computational effort on larger scale problems.

3.1. Convergence rate

Via the RIP-constant δ_s theorem 2.7 gives a direct dependence of the convergence rate on the sparsity of the solution and the properties of the matrix, whereas theorem 2.11 is harder to interpret: it is straight-forward to deduce the existence of parameter regimes in which linear convergence occurs but hard to quantify the rate in terms of the parameters. While numerical evidence for linear convergence of the infimal convolution formulation is observed in section 3.2, we continue by validating theorem 2.7 in two experiments. In both, we take $q = 1/2$, and add pre- and post-measurement Gaussian noise terms, $\mathbf{v}$ and $\boldsymbol{\xi}$, with noise level $\frac{\|\mathbf{v}\|_2}{\|\mathbf{u}^\dagger\|_2} = \frac{\|\boldsymbol{\xi}\|_2}{\|\mathbf{u}^\dagger\|_2} = 0.1$. We choose an admissible α according to remark 2.8 and tune it such that the reconstructed



(a) Varying β. (b) Varying m.

Figure 1. In the left panel we consider $\mathbf{A} \in \mathbb{R}^{200\times 600}$ and vary the parameter β, whereas in the right panel we consider $\mathbf{A} \in \mathbb{R}^{m\times 600}$ and vary the number of measurements $m \in \{100, 200, 300, 400\}$.

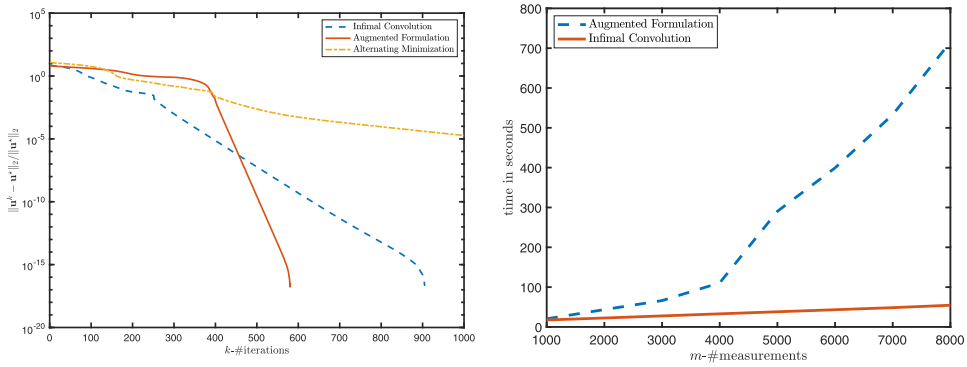
signal shares its support size with the ground-truth. Both illustrations in figure 1 plot the relative error between the iterates $\mathbf{u}^k$ and the stationary point $\mathbf{u}^\star$ against the number of proximal gradient descent steps.

Varying the penalty parameter. In the first experiment we take a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{200\times 600}$, a 20-sparse signal $\mathbf{u}^\dagger$, and vary β. Theorem 2.7 predicts that smaller values of β allow to take larger stepsizes, though the convergence constants are (essentially) the same. This effect is readily observed in figure 1(a). Note that we can also observe that for smaller β the algorithm reaches the steep part of the curve faster. This is due to the fact that the convergence of iterates is initially slow (until the support is identified) and larger step-sizes allow to reduce the support size faster. The overall speed-up allowed by a smaller β can be by up to a two-fold, in terms of the number of iterations needed to reach the desired accuracy level.

Varying the measurements. In the second experiment we consider a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{m\times 600}$, for $m \in \{100, 200, 300, 400\}$, and a 20-sparse signal $\mathbf{u}^\dagger$. Varying the number of measurements changes the RIP of the measurement matrix (a larger m decreases δ_s, see remark 2.3), and per theorem 2.7 should affect the convergence constant. Figure 1(b) shows exactly that. An analogous effect can be observed for different classes of measurement matrices, such as partial Toeplitz, or partial circulant matrices with Rademacher or Gaussian entries, but those results have not been included for the sake of brevity.

3.2. Computational comparison

Iteration count. In order to provide numerical evidence for our initial statement that alternating minimization is highly sub-optimal, in figure 2(a) we look at the decay of the relative error over the number of basic iterations, i.e. the number of thresholded gradient descent steps, of all three discussed approaches: alternating minimization (3), augmented formulation (9), and infimal convolution (15). In this experiment, we use a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{100\times 500}$, the original signal is 14-sparse, $q = 1/2$ and the parameter α, β, and μ are selected so that each method returns a 13-sparse vector. The x-axis refers to the number of times the proximal operator is called while the y-axis shows the relative error. The considerably worse performance of alternating minimization is due to the fact that it requires (too) many thresholded gradient steps to solve, for each $k \in \mathbb{N}$, sub-problems for the $\mathbf{u}^k$ component up to pre-fixed accuracy $\varepsilon = 10^{-8}$. Thus, the algorithm performs hardly any alternating steps.



(a) Convergence Rate (b) Running Time (50 iterations)

Figure 2. In the left panel we look at the relative error with respect to the number of times the proximal operator is called for $\mathbf{A} \in \mathbb{R}^{100\times 500}$ and $\mathbf{u}^\dagger \in \mathbb{R}^{500}$ is 14-sparse. In the right panel we compare average running time of augmented and infimal convolution formulations when reconstructing a 100-sparse signal $\mathbf{u}^\dagger \in \mathbb{R}^{5000}$ from m measurements, that vary from 1000 to 8000.

Computation time. To now illustrate the differences between augmented and infimal convolution formulation in terms of computational complexity, we perform the following experiment. We set the parameters generically to $\alpha = 0.02$, $\beta = 0.2$, and $\mu = 0.1$, and reconstruct a 100-sparse signal $\mathbf{u}^\dagger \in \mathbb{R}^{5000}$ from measurements $\mathbf{y} \in \mathbb{R}^m$, for m varying from 1000 (sub-sampling) to 8000 (over-sampling). We again take $q = 1/2$, and add pre- and post-measurement noise terms, $\mathbf{v}$ and $\boldsymbol{\xi}$, with noise level 0.1. Averaging over 20 random realizations of $\mathbf{u}^\dagger$, we record for augmented (9) and infimal convolution approach (15) the time needed to perform 50 iterations. After only this many iterations none of the two algorithms has converged, though this already suffices to make a point regarding the computational cost since both algorithms incur the same cost (i.e. the gap remains the same) in the remaining iterations. As figure 2(b) shows, the additional computation of $\mathbf{B}_\beta$ in (9) causes a massive additional workload leading to limited applicability of the augmented approach in large-scale settings. In contrast, the infimal convolution formulation is hardly affected by the increase in the number of measurements. Though the augmented approach tends to converge in fewer iterations, cf figure 2(a), the additional iterations needed by the infimal convolution formulation to reach a comparable level of accuracy do not close the gap in computation time. Note that we do not include alternating minimization here since it requires many more iterations (in the sense of single thresholded gradient descent steps) to show similar reconstruction performance as both proximal descents, and hence could not compete with those two algorithms.

4. Discussion

In the present work we discussed the benefits of multi-penalty regularization for support recovery of signals when pre-measurement noise is amplified by the measurement operator and numerical challenges in solving the corresponding variational formulation. Since alternating minimization is for this task sub-optimal in terms of both the computational efficiency and theoretical analysis, we proposed a novel reduction to single-penalty regularization based on infimal convolution, and compared this new approach to an existing reduction based on augmented formulations. Moreover, we established linear convergence for both single-penalty reductions

and showed that our new approach omits a computational bottleneck that is unavoidable in the augmented approach, and causes a significant additional computational workload if the number of measurements increases. There are several interesting open questions left for future work.

First, in remark 2.13 we observed, for $q = 1$, a connection between the infimal convolution formulation and the proximal descent on the ℓ_2-norm of the gradient of a Moreau-regularized ℓ_1-functional. As we have not seen a comparable relation in the context of multi-penalty regularization so far, we are curious whether this observation can be extended to the case $0 < q < 1$. If so, this might provide valuable insights into non-convex optimization.

Second, as the reader might have noticed, great parts of the arguments we used (support stabilization, sign stabilization, etc.) are not restricted to finite dimensions. In light of more general settings of multi-penalty regularization in [21] and single-penalty regularization in [9], it would be fruitful to transfer our findings to general separable Hilbert spaces as well.

Third, we mention that when using the infimal convolution based approach, in some experiments it was possible to choose μ much larger than suggested by theorem 2.11, while still observing reliable convergence of the program. We wonder whether there is an alternative proof leading to a relaxed condition on μ resembling the assumption in theorem 2.7.

Let us conclude by emphasizing that the infimal convolution formulation can as well be applied if regularizers other than the ℓ_q-norm are used in the multi-penalty problem, e.g. smoothly clipped absolute deviation [13], minimax concave penalty [29], and log-sum penalty [10]. In those cases the more general single-penalty rate analysis in [28] should prove useful as a tool.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ZK and VN acknowledge the support from RCN-funded FunDaHD Project No. 251149/O70. JM acknowledges the support of DFG-SPP 1798.

Appendix A. Proofs

A.1. Proof of lemma 2.4

For a fixed $\mathbf{u}$ the minimization of $\mathcal{T}^q_{\alpha,\beta}$ in (4) with respect to $\mathbf{v}$ reduces to Tikhonov minimization, and thus the solution satisfies

$$\mathbf{v} = v(\mathbf{u}) = \left(\beta\,\mathbf{Id}_n + \mathbf{A}^\top\mathbf{A}\right)^{-1}\left(\mathbf{A}^\top\mathbf{y} - \mathbf{A}^\top\mathbf{A}\mathbf{u}\right). \tag{19}$$

Rewriting the above expression we have

$$\beta v(\mathbf{u}) = \mathbf{A}^\top(\mathbf{y} - \mathbf{A}\mathbf{u}) - \mathbf{A}^\top\mathbf{A}v(\mathbf{u}).$$

Plugging this expression into (4) the minimization problem for u is rewritten as

$$\mathcal{T}^q_{\alpha,\beta}(\mathbf{u}, v(\mathbf{u})) = \frac{1}{2}\left\langle \mathbf{A}(\mathbf{u} + v(\mathbf{u})) - \mathbf{y}, \mathbf{A}\mathbf{u} - \mathbf{y}\right\rangle + \frac{\alpha}{q}\|\mathbf{u}\|_q^q.$$

The Woodbury identity for invertible matrices $\mathbf{V} \in \mathbb{R}^{m\times m}$, $\mathbf{W} \in \mathbb{R}^{n\times n}$ and matrices $\mathbf{M}_1 \in \mathbb{R}^{m\times n}$, $\mathbf{M}_2 \in \mathbb{R}^{n\times m}$ reads

$$\left(\mathbf{V}+\mathbf{M}_1\mathbf{W}^{-1}\mathbf{M}_2\right)^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{M}_1\left(\mathbf{W}+\mathbf{M}_2\mathbf{V}^{-1}\mathbf{M}_1\right)^{-1}\mathbf{M}_2\mathbf{V}^{-1}. \tag{20}$$

Using (19), this gives

$$\begin{aligned}\mathbf{A}(\mathbf{u}+v(\mathbf{u}))-\mathbf{y} &= \mathbf{A}v(\mathbf{u})+\mathbf{A}\mathbf{u}-\mathbf{y}\\ &= \left(\mathbf{Id}_m - \mathbf{A}\left(\beta\mathbf{Id}_n+\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\right)(\mathbf{A}\mathbf{u}-\mathbf{y})\\ &= \left(\mathbf{Id}_m+\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1}(\mathbf{A}\mathbf{u}-\mathbf{y}).\end{aligned}$$

Plugging this expression back into $\mathcal{T}^q_{\alpha,\beta}(\mathbf{u}, v(\mathbf{u}))$, and extracting the square root, we have $\mathcal{T}^q_{\alpha,\beta}(\mathbf{u}, v(\mathbf{u})) = \mathcal{F}_\beta(\mathbf{u})$. Minimizing over $\mathbf{u}$ and using the following simple observation gives the conclusion.

Lemma A.1. *If $\mathbf{u}^q_{\alpha,\beta}$ is a local minimizer of (7), then the pair $(\mathbf{u}^q_{\alpha,\beta}, v(\mathbf{u}^q_{\alpha,\beta}))$ with $v(u)$ defined in (6), is a local minimizer of $\mathcal{T}^q_{\alpha,\beta}$ in (4).*

Proof. Let $\mathbf{u}^q_{\alpha,\beta}$ be a local minimizer of (7) and assume there exists a sequence $(\mathbf{u}^k, \mathbf{v}^k) \to (\mathbf{u}^q_{\alpha,\beta}, v(\mathbf{u}^q_{\alpha,\beta}))$ such that $\mathcal{T}^q_{\alpha,\beta}(\mathbf{u}^k, \mathbf{v}^k) < \mathcal{T}^q_{\alpha,\beta}(\mathbf{u}^q_{\alpha,\beta}, v(\mathbf{u}^q_{\alpha,\beta}))$, for all $k \in \mathbb{N}$. We then have

$$\mathcal{F}_\beta(\mathbf{u}^k) = \mathcal{T}^q_{\alpha,\beta}(\mathbf{u}^k, v(\mathbf{u}^k)) \leqslant \mathcal{T}^q_{\alpha,\beta}(\mathbf{u}^k, \mathbf{v}^k) < \mathcal{T}^q_{\alpha,\beta}(\mathbf{u}^q_{\alpha,\beta}, v(\mathbf{u}^q_{\alpha,\beta})) = \mathcal{F}_\beta(\mathbf{u}^q_{\alpha,\beta}),$$

where the first inequality follows from the minimality of $v(\mathbf{u}^k)$. This contradicts the assumption that $\mathbf{u}^q_{\alpha,\beta}$ is a local minimizer of (7). □

A.2. Proof of lemma 2.6

First note that

$$\begin{aligned}&\operatorname{prox}_{\mu,\frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{u}^k - \mu\mathbf{B}_\beta^\top(\mathbf{B}_\beta\mathbf{u}^k - \mathbf{y}_\beta))\\ &\quad= \operatorname{prox}_{\mu,\frac{\alpha}{q}\|\cdot\|_q^q}\left(\mathbf{u}^k - \mu\mathbf{A}^\top\left(\mathbf{Id}_m+\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1}\left(\mathbf{A}\mathbf{u}^k-\mathbf{y}\right)\right)\end{aligned}$$

while

$$\begin{aligned}&\operatorname{prox}_{\mu,\frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{u}^k - \mu\mathbf{A}^\top(\mathbf{A}\mathbf{u}^k + \mathbf{A}v(\mathbf{u}^k) - \mathbf{y})) =\\ &\quad\operatorname{prox}_{\mu,\frac{\alpha}{q}\|\cdot\|_q^q}\left(\mathbf{u}^k - \mu\left(\mathbf{A}^\top - \mathbf{A}^\top\mathbf{A}(\beta\mathbf{Id}_n+\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\right)\left(\mathbf{A}\mathbf{u}^k-\mathbf{y}\right)\right).\end{aligned}$$

Hence, it suffices to show that

$$\mathbf{A}^\top\left(\mathbf{Id}_m+\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1} = \mathbf{A}^\top - \mathbf{A}^\top\mathbf{A}(\beta\,\mathbf{Id}_n+\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top.$$

Extracting $\mathbf{A}^\top$ from the left and using the Woodbury identity (20) with $\mathbf{M}_1 = \mathbf{A}, \mathbf{M}_2 = \mathbf{A}^\top$, $\mathbf{W} = \beta\mathbf{Id}_n$, and $\mathbf{V} = \mathbf{Id}_m$ the conclusion follows.

A.3. Proof of theorem 2.7

In order to prove theorem 2.7, we have to control the eigenvalues of $\mathbf{B}_\beta^\top \mathbf{B}_\beta$ characterizing the growth of the data fidelity term in (7).

Lemma A.2. *For $\mathbf{B}_\beta \in \mathbb{R}^{m\times n}$ defined as in lemma 2.4,*

$$L := \|\mathbf{B}_\beta^\top \mathbf{B}_\beta\| = \left(\|\mathbf{A}\|^{-2} + \beta^{-1}\right)^{-1},$$

is the Lipschitz-constant of the gradient of the augmented data-fidelity term $\frac{1}{2}\|\mathbf{B}_\beta \mathbf{u} - \mathbf{y}_\beta\|_2^2$. Moreover, for any $I \subset [n]$,

$$\lambda_{min}(\mathbf{B}_{\beta,I}^\top \mathbf{B}_{\beta,I}) \geqslant \left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1} \lambda_{min}(\mathbf{A}_I^\top \mathbf{A}_I).$$

Proof. Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ denote the SVD of $\mathbf{A}$.

This gives

$$\mathbf{B}_\beta^\top \mathbf{B}_\beta = \mathbf{V}\boldsymbol{\Sigma}^\top \left(\mathbf{Id}_m + \frac{\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top}{\beta}\right)^{-1} \boldsymbol{\Sigma}\mathbf{V}^\top, \tag{21}$$

so that $\|\mathbf{B}_\beta^\top \mathbf{B}_\beta\| = \left(\|\mathbf{A}\|^{-2} + \beta^{-1}\right)^{-1}$.

By (21), we have for any $\mathbf{z} \in \mathbb{R}^n$

$$\begin{aligned}\left|\mathbf{z}^\top \mathbf{B}_\beta^\top \mathbf{B}_\beta \mathbf{z}\right| &= \left|\mathbf{z}^\top \mathbf{V}\boldsymbol{\Sigma}^\top \left(\mathbf{Id}_m + \frac{\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top}{\beta}\right)^{-1} \boldsymbol{\Sigma}\mathbf{V}^\top \mathbf{z}\right| \geqslant \left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1} \left|\mathbf{z}^\top \mathbf{V}\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}\mathbf{V}^\top \mathbf{z}\right| \\ &= \left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1} \left|\mathbf{z}^\top \mathbf{A}^\top \mathbf{A}\mathbf{z}\right|, \end{aligned} \tag{22}$$

implying the second claim. □

We can now show that all, up to finitely many, iterates $\left(\mathbf{u}^k\right)_{k=1}^\infty$ generated by (9) share the same support and sign pattern. The proof is standard and follows [9].

Lemma A.3 (Support and sign recovery). *Assume $\beta > 0$, $0 < q \leqslant 1$, and $\mu < \|\mathbf{A}\|^{-2} + \beta^{-1}$. Then the iterates $\left(\mathbf{u}^k\right)_{k=1}^\infty$ satisfy $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \to 0$ as $k \to \infty$. Moreover, all iterates, up to finitely many, have the same support and sign pattern.*

Proof. Since $\mu < \|\mathbf{A}\|^{-2} + \beta^{-1} = \frac{1}{L}$ we have $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \to 0$ as $k \to \infty$ by [9, corollary 2.1]. Now, since the range of $\mathrm{prox}_{\mu,\lambda\psi}$ is $(-\infty, -\lambda_{\mu,q}] \cup \{0\} \cup [\lambda_{\mu,q}, \infty)$, it follows that the the absolute value of a non-zero entry of $\mathbf{u}^k$, for $k \geqslant 1$, is at least $\lambda_{\mu,q}$. Thus, if $\mathrm{supp}(\mathbf{u}^{k+1}) \neq \mathrm{supp}(\mathbf{u}^k)$ we have $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \geqslant \lambda_{\mu,q}$, and analogously, if $\mathrm{sgn}(\mathbf{u}^{k+1}) \neq \mathrm{sgn}(\mathbf{u}^k)$ we have $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \geqslant 2\lambda_{\mu,q}$. Thus, since $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \to 0$ as $k \to \infty$, sign and support can change only finitely many times. □

Proof of theorem 2.7. By lemma A.3 there exists k_0 such that for all $k \geqslant k_0$ the support of $\mathbf{u}^k$ is finite, and support and sign of $\mathbf{u}^k$ is equal to that of $\mathbf{u}^\star$. Thus, by [9, proposition 2.3], $\mathbf{u}^\star$ is a fixed point of (9). Denote $I = \mathrm{supp}(\mathbf{u}^\star)$ with $|I| \leqslant s$. The definition of proximal operator in (10) and the Karush–Kuhn–Tucker (KKT) conditions yield

$$\alpha\ \mathrm{sgn}(\mathsf{u}_i^\star)\left|\mathsf{u}_i^\star\right|^{q-1} = -(\mathbf{B}_\beta^\top(\mathbf{B}_\beta \mathbf{u}^\star - \mathbf{y}_\beta))_i, \quad i \in I,$$

and

$$\mathsf{u}_i^{k+1} + \alpha\mu\ \mathrm{sgn}(\mathsf{u}_i^{k+1})\,|\mathsf{u}_i^{k+1}|^{q-1} = \mathsf{u}_i^k - \mu(\mathbf{B}_\beta^\top(\mathbf{B}_\beta\mathbf{u}^k - \mathbf{y}_\beta))_i, \quad i \in I.$$

Subtracting the two equations on the index set I, and denoting $\psi(\mathbf{u}) = \frac{1}{q}\|\mathbf{u}\|_q^q$, we have

$$\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star + \alpha\mu\left(\psi'(\mathbf{u}_I^{k+1}) - \psi'(\mathbf{u}_I^\star)\right) = \mathbf{u}_I^k - \mathbf{u}_I^\star - \mu\big(\mathbf{B}_\beta^\top\mathbf{B}_\beta(\mathbf{u}^k - \mathbf{u}^\star)\big)_I, \quad (23)$$

where $\psi'(\mathbf{u}) = (\mathrm{sgn}(\mathsf{u}_i)|\mathsf{u}_i|^{q-1})_{i\in[n]}$ is acting entry-wise. Note that since $k \geqslant k_0$ we have $\mathrm{sgn}(\mathbf{u}_I^\star) = \mathrm{sgn}(\mathbf{u}_I^{k+1})$ and $\left\|\mathbf{u}^\star - \mathbf{u}^k\right\|_2 = \|\mathbf{u}_I^\star - \mathbf{u}_I^k\|_2$. A straightforward calculation gives

$$\mathbf{u}_I^k - \mathbf{u}_I^\star - \mu\big(\mathbf{B}_\beta^\top\mathbf{B}_\beta(\mathbf{u}^k - \mathbf{u}^\star)\big)_I = (\mathbf{Id}_s - \mu\mathbf{M}_{I,I})\left(\mathbf{u}_I^k - \mathbf{u}_I^\star\right)$$

where $\mathbf{M} = \mathbf{B}_\beta^\top\mathbf{B}_\beta$. Taking the inner product of (23) with $\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star$, and applying the Cauchy–Schwartz inequality, we get

$$\begin{aligned}\|\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star\|_2^2 - \alpha\mu\left\langle\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star, \psi'(\mathbf{u}_I^{k+1}) - \psi'(\mathbf{u}_I^\star)\right\rangle \\ \leqslant \|\mathbf{Id}_s - \mu\mathbf{M}_{I,I}\|\|\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star\|_2\|\mathbf{u}_I^k - \mathbf{u}_I^\star\|_2.\end{aligned}$$

Since ψ is twice differentiable, and $\mathbf{u}^{k+1}$ and $\mathbf{u}^\star$ have the same sign and support, we have for the second term

$$\begin{aligned}\left\langle\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star, \psi'(\mathbf{u}_I^{k+1}) - \psi'(\mathbf{u}_I^\star)\right\rangle &= \sum_{i\in I}(\mathsf{u}_i^{k+1} - \mathsf{u}_i^\star)\left(\psi'(\mathsf{u}_i^{k+1}) - \psi'(\mathsf{u}_i^\star)\right) \\ &= \sum_{i\in I}\psi''(C_i^{k+1})(\mathsf{u}_i^{k+1} - \mathsf{u}_i^\star)^2,\end{aligned}$$

where C_i^{k+1} lies between u_i^{k+1} and $\mathsf{u}_i^\star$, and $\psi''(\mathsf{u}) = (q-1)\mathsf{u}^{q-2}$. Since $\mathbf{u}^k \to \mathbf{u}^\star$, we may assume k_0 sufficiently large to guarantee $\mathsf{u}_i^k \geqslant \frac{1}{2}\mathsf{u}_i^\star$, for all $k \geqslant k_0$ and $i \in I$. Consequently,

$$\left|\psi''(C_i^{k+1})\right| = |q-1|\,|C_i^{k+1}|^{q-2} \leqslant (1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}.$$

Thus,

$$\begin{aligned}\|\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star\|_2^2 - \alpha\mu\left\langle\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star, \psi'(\mathbf{u}_I^{k+1}) - \psi'(\mathbf{u}_I^\star)\right\rangle \\ \geqslant \left(1 - \mu\alpha(1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}\right)\|\mathbf{u}_I^{k+1} - \mathbf{u}_I^\star\|_2^2.\end{aligned}$$

On the other hand, since $\mu < (\lambda_{\max}(\mathbf{M}))^{-1} \leqslant (\lambda_{\min}(\mathbf{M}_{I,I}))^{-1}$, we have

$$\|\mathbf{Id}_s - \mu\mathbf{M}_{I,I}\| = 1 - \mu\lambda_{\min}(\mathbf{M}_{I,I}) \leqslant 1 - \mu\left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1}\lambda_{\min}(\mathbf{A}_I^\top\mathbf{A}_I),$$

by lemma A.2. Thus,

$$\|\mathbf{u}^{k+1} - \mathbf{u}^\star\|_2 \leqslant \frac{1 - \mu\left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right)^{-1}\lambda_{\min}(\mathbf{A}_I^\top\mathbf{A}_I)}{1 - \mu\alpha(1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}}\|\mathbf{u}^k - \mathbf{u}^\star\|_2.$$

Together with the RIP of $\mathbf{A}$ this yields the claim. □

A.4. Proof of lemma 2.10

Let $\mathbf{x} \in \mathbb{R}^n$ be fixed and assume $f(0) = 0$ without loss of generality. We have

$$\begin{aligned}\operatorname{prox}_{\mu,\lambda M_{t,f}}(\mathbf{x}) &= \operatorname*{arg\,min}_{\mathbf{z}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + \mu\lambda M_{t,f}(\mathbf{z}) \\ &= \operatorname*{arg\,min}_{\mathbf{z}\in\mathbb{R}^n} \inf_{\tilde{\mathbf{z}}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + \mu\lambda f(\tilde{\mathbf{z}}) + \frac{\mu\lambda}{2t}\|\mathbf{z}-\tilde{\mathbf{z}}\|_2^2 \\ &= \operatorname*{arg\,min}_{\mathbf{z}\in\mathbb{R}^n} \inf_{\tilde{\mathbf{z}}\in\mathbb{R}^n} h(\mathbf{z},\tilde{\mathbf{z}}).\end{aligned}$$

By f being lower semi-continuous and bounded from below, we have

$$\inf_{\mathbf{z},\tilde{\mathbf{z}}\in\mathbb{R}^n} h(\mathbf{z},\tilde{\mathbf{z}}) = \min_{\mathbf{z},\tilde{\mathbf{z}}\in\mathbb{R}^n} h(\mathbf{z},\tilde{\mathbf{z}}),$$

implying $\operatorname{prox}_{\mu,\lambda M_{t,f}}(\mathbf{x}) \neq \emptyset$. Denote by $\mathcal{E}_{\tilde{\mathbf{z}}} = \{\theta\mathbf{x} + (1-\theta)\tilde{\mathbf{z}} : \theta \in [0,1]\}$ the line connecting $\mathbf{x}$ and $\tilde{\mathbf{z}}$. Since $\mathcal{E}_{\tilde{\mathbf{z}}}$ is convex, we have $h(\mathbf{P}_{\mathcal{E}_{\tilde{\mathbf{z}}}}(\mathbf{z}),\tilde{\mathbf{z}}) \leqslant h(\mathbf{z},\tilde{\mathbf{z}})$, for any $\mathbf{z},\tilde{\mathbf{z}} \in \mathbb{R}^n$, with equality if and only if $\mathbf{P}_{\mathcal{E}_{\tilde{\mathbf{z}}}}(\mathbf{z}) = \mathbf{z}$. Consequently, if $(\mathbf{z},\tilde{\mathbf{z}})$ solves the above program, we have $\mathbf{z} = \theta\mathbf{x} + (1-\theta)\tilde{\mathbf{z}}$ for some $\theta \in [0,1]$. Let us define

$$\tilde{h}(\theta,\tilde{\mathbf{z}}) = h(\theta\mathbf{x} + (1-\theta)\tilde{\mathbf{z}},\tilde{\mathbf{z}}) = \left(\frac{(1-\theta)^2}{2} + \frac{\mu\lambda\theta^2}{2t}\right)\|\mathbf{x}-\tilde{\mathbf{z}}\|_2^2 + \mu\lambda f(\tilde{\mathbf{z}}).$$

By the above considerations we have

$$\min_{\mathbf{z},\tilde{\mathbf{z}}\in\mathbb{R}^n} h(\mathbf{z},\tilde{\mathbf{z}}) = \min_{\tilde{\mathbf{z}}\in\mathbb{R}^n} \min_{\theta\in[0,1]} \tilde{h}(\theta,\tilde{\mathbf{z}}),$$

where there is a one-to-one correspondence between solutions $(\mathbf{z}^\star,\tilde{\mathbf{z}}^\star)$ of the left side and solutions $(\theta^\star,\tilde{\mathbf{z}}^\star)$. Moreover, it follows easily that for $\tilde{\mathbf{z}} \in \mathbb{R}^n$ fixed,

$$\theta^\star = \operatorname*{arg\,min}_{\theta\in[0,1]} \tilde{h}(\theta,\tilde{\mathbf{z}}) = \frac{1}{1+\frac{\mu\lambda}{t}},$$

which is independent of $\tilde{\mathbf{z}}$. Thus, the claim follows since

$$\operatorname*{arg\,min}_{\tilde{\mathbf{z}}\in\mathbb{R}^n} \tilde{h}(\theta^\star,\tilde{\mathbf{z}}) = \operatorname*{arg\,min}_{\tilde{\mathbf{z}}\in\mathbb{R}^n} \frac{\mu\lambda}{t}\frac{1}{1+\frac{\mu\lambda}{t}}\frac{\|\mathbf{x}-\tilde{\mathbf{z}}\|_2^2}{2} + \mu\lambda f(\tilde{\mathbf{z}}) = \operatorname{prox}_{(t+\mu\lambda),f}(\mathbf{x}).$$

A.5. Proof of theorem 2.11

As in the proof of theorem 2.7, the first step is to control support and signs of the iterates. Recall that, for $\mathbf{w}^k$ as in (15), we denote by $\mathbf{u}^k = \operatorname{prox}_{\frac{1}{\beta},\frac{\alpha}{q}\|\cdot\|_q^q}(\mathbf{w}^k)$ the sequence of minimizers attaining $g(\mathbf{w}^k)$, by $\mathbf{v}^k = \mathbf{w}^k - \mathbf{u}^k$, and that by (16) we have

$$\begin{cases}\mathbf{w}^k = \operatorname*{arg\,min}_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2\mu}\|\mathbf{w}-\mathbf{w}^{k-1}+\mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^{k-1}-\mathbf{y})\|_2^2 + \frac{\beta}{2}\|\mathbf{w}-\mathbf{u}^k\|_2^2 \\ \mathbf{u}^k = \operatorname*{arg\,min}_{\mathbf{u}\in\mathbb{R}^n} \frac{\beta}{2}\|\mathbf{u}-\mathbf{w}^k\|_2^2 + \frac{\alpha}{q}\|\mathbf{u}\|_q^q.\end{cases} \tag{24}$$

Lemma A.4 (Sign and support stability). *Assume $\mu < \|\mathbf{A}\|^{-2}$. Then the successive iterates $\|\mathbf{w}^{k+1} - \mathbf{w}^k\|_2$, $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2$, and $\|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2$ converge to zero and all but finitely many iterates $\mathbf{u}^k$ share the same finite support and the same signs.*

Proof. First, note that g is a proper and coercive function. Second, as $g(\mathbf{w}) = \inf_{\mathbf{u}\in\mathbb{R}^n} f(\mathbf{u}, \mathbf{w})$, for f continuous, we obtain continuity of g at any point $\mathbf{w} \in \mathbb{R}^n$ since by coercivity of f the infimum can be restricted to a finite ball and the infimum of continuous functions on a compact set is continuous. Consequently, by [9, corollary 2.1] and the assumption on μ we have $\|\mathbf{w}^{k+1} - \mathbf{w}^k\|_2 \to 0$, for $\mathbf{w}^{k+1} = \mathrm{prox}_{\mu,g}(\mathbf{w}^k - \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y}))$. By the KKT-conditions of (24), we obtain

$$0 = (\mathbf{w}^{k+1} - \mathbf{w}^k) + \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y}) + \beta\mu\mathbf{v}^{k+1},$$
$$0 = (\mathbf{w}^{k} - \mathbf{w}^{k-1}) + \mu\mathbf{A}^\top(\mathbf{A}\mathbf{w}^{k-1} - \mathbf{y}) + \beta\mu\mathbf{v}^{k}.$$

Subtracting the two equations gives $\|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2 \to 0$, and $\mathbf{u}^k = \mathbf{w}^k - \mathbf{v}^k$ yields $\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 \to 0$. The second claim follows as in lemma A.3, since $\mathbf{u}^k$ is a thresholded version of $\mathbf{w}^k$. □

Proof of theorem 2.11. First note that $\mathbf{w}^k \to \mathbf{w}^\star$ implies via lemma A.4 that $\mathbf{u}^k \to \mathbf{u}^\star$ and $\mathbf{v}^k \to \mathbf{v}^\star$. Furthermore, $\mathbf{w}^\star$ is a fixed point of (15), by [9, proposition 2.3]. By lemma A.4 there exists k_0 such that for all $k \geqslant k_0$ the support of $\mathbf{u}^k$ is finite, and support and sign of $\mathbf{u}^k$ is equal to that of $\mathbf{u}^\star$. Denote $I = \mathrm{supp}(\mathbf{u}^\star)$. By the KKT-conditions of (24), we get

$$i \in I : \begin{cases} \alpha\mu\,\mathrm{sign}(\mathsf{u}_i^\star)|\mathsf{u}_i^\star|^{q-1} = -\mu(\mathbf{A}^\top(\mathbf{A}\mathbf{w}^\star - \mathbf{y}))_i, \\ \mathsf{w}_i^{k+1} + \alpha\mu\,\mathrm{sign}(\mathsf{u}_i^{k+1})|\mathsf{u}_i^{k+1}|^{q-1} = \mathsf{w}_i^k - \mu(\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y}))_i, \end{cases}$$

and

$$i \notin I : \begin{cases} 0 = \beta\mu\mathsf{w}_i^\star + \mu(\mathbf{A}^\top(\mathbf{A}\mathbf{w}^\star - \mathbf{y}))_i, \\ 0 = (1+\beta\mu)\mathsf{w}_i^{k+1} - \mathsf{w}_i^k + \mu(\mathbf{A}^\top(\mathbf{A}\mathbf{w}^k - \mathbf{y}))_i, \end{cases}.$$

For $\psi(\mathbf{u}) = \frac{1}{q}\|\mathbf{u}\|_q^q$ with $\psi'(\mathbf{u}) = (\mathrm{sgn}(\mathsf{u}_i)|\mathsf{u}_i|^{q-1})_{i\in[n]}$ acting entry-wise, this implies

$$(\mathbf{w}^{k+1} - \mathbf{w}^\star)_I + \alpha\mu(\psi'(\mathbf{u}^{k+1}) - \psi'(\mathbf{u}^\star)) = (\mathbf{w}^k - \mathbf{w}^\star)_I - \mu\mathbf{A}_I^\top\mathbf{A}(\mathbf{w}^k - \mathbf{w}^\star) \tag{25}$$

and

$$(1+\mu\beta)(\mathbf{w}^{k+1} - \mathbf{w}^\star)_{I^c} = (\mathbf{w}^k - \mathbf{w}^\star)_{I^c} - \mu\mathbf{A}_{I^c}^\top\mathbf{A}(\mathbf{w}^k - \mathbf{w}^\star). \tag{26}$$

Repeating the steps as in theorem 2.7, from (25) we get

$$\left(1 - \alpha\mu(1-q)\left(\frac{d_{\min}}{2}\right)^{q-2}\right)\|(\mathbf{w}^{k+1} - \mathbf{w}^\star)_I\|_2^2 \leqslant \|\mathbf{P}_I - \mu\mathbf{A}_I^\top\mathbf{A}\|\|\mathbf{w}^k - \mathbf{w}^\star\|_2\|(\mathbf{w}^{k+1} - \mathbf{w}^\star)_I\|_2$$

and from (26) we obtain

$$(1+\mu\beta)\|(\mathbf{w}^{k+1} - \mathbf{w}^\star)_{I^c}\|_2 \leqslant \|\mathbf{P}_{I^c} - \mu\mathbf{A}_{I^c}^\top\mathbf{A}\|\|\mathbf{w}^k - \mathbf{w}^\star\|_2.$$

Squaring and summing the last two equations, the claim follows by orthogonality of $(\mathbf{w}^{k+1} - \mathbf{w}^\star)_I$ and $(\mathbf{w}^{k+1} - \mathbf{w}^\star)_{I^c}$. □

Appendix B. Coherence bound

The following lemma bounds the coherence of $\mathbf{B}_\beta$ in terms of the coherence of $\mathbf{A}$. The bound becomes tight for large choices of β.

Lemma B.1. *We have*

$$\operatorname{coh}(\mathbf{B}_\beta) \leqslant \left(1 + \frac{\|\mathbf{A}\|^2}{\beta}\right) \operatorname{coh}(\mathbf{A}) + \frac{\|\mathbf{A}\|^2}{\beta}.$$

Proof. Recall that the coherence of a matrix is defined as

$$\operatorname{coh}(\mathbf{M}) = \max_{i \neq j} \frac{\left|\mathbf{m}_i^\top \mathbf{m}_j\right|}{\|\mathbf{m}_i\|_2 \|\mathbf{m}_j\|_2},$$

where $\mathbf{m}_i$ is the ith column of $\mathbf{M}$. Define $\mathbf{Q}_\beta = \mathbf{Id}_m + \frac{\mathbf{A}\mathbf{A}^\top}{\beta}$, so that $\mathbf{B}_\beta = \mathbf{Q}_\beta^{-1/2}\mathbf{A}$, and let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the SVD of $\mathbf{A}$. This gives

$$\mathbf{Q}_\beta^{-1} - \mathbf{Id}_m = \left(\mathbf{Id}_m + \frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1} - \mathbf{Id}_m = \mathbf{U}\left(\left(\mathbf{Id}_m + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^\top}{\beta}\right)^{-1} - \mathbf{Id}_m\right)\mathbf{U}^\top.$$

Therefore,

$$\left\|\mathbf{Q}_\beta^{-1} - \mathbf{Id}_m\right\| = \left\|\left(\mathbf{Id}_m + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^\top}{\beta}\right)^{-1} - \mathbf{Id}_m\right\| = \frac{c_\beta}{1 + c_\beta},$$

for $c_\beta = \frac{\|\mathbf{A}\|^2}{\beta}$, and by triangle inequality and Cauchy–Schwarz

$$|\mathbf{b}_i^\top \mathbf{b}_j| = |\mathbf{a}_i^\top \mathbf{Q}_\beta^{-1} \mathbf{a}_j| \leqslant |\mathbf{a}_i^\top \mathbf{a}_j| + \frac{c_\beta}{1 + c_\beta} \|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2,$$

for all columns $\mathbf{b}_i, \mathbf{b}_j$ of $\mathbf{B}_\beta$. By the same argument we compute

$$\mathbf{Q}_\beta^{-1/2} - \mathbf{Id}_m = \mathbf{U}\left(\left(\mathbf{Id}_m + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^\top}{\beta}\right)^{-1/2} - \mathbf{Id}_m\right)\mathbf{U}^\top,$$

giving

$$\|\mathbf{Q}_\beta^{-1/2} - \mathbf{Id}_m\| = 1 - \sqrt{\frac{\beta}{\|\mathbf{A}\|^2 + \beta}} = 1 - (c_\beta + 1)^{-1/2}.$$

This yields

$$\|\mathbf{b}_i\|_2 \geqslant \|\mathbf{a}_i\|_2 - \|(\mathbf{Q}_\beta^{-1/2} - \mathbf{Id}_m)\mathbf{a}_i\|_2 \geqslant (c_\beta + 1)^{-1/2} \|\mathbf{a}_i\|_2 \tag{27}$$

which implies

$$\operatorname{coh}(\mathbf{B}_\beta) = \max_{i\neq j} \frac{\left|\mathbf{b}_i^\top \mathbf{b}_j\right|}{\|\mathbf{b}_i\|_2\|\mathbf{b}_j\|_2} \leqslant (1+c_\beta)\left(\operatorname{coh}(\mathbf{A}) + \frac{c_\beta}{1+c_\beta}\right).$$

□

For small β, the bound in lemma B.1 is lossy. However, we can show that the coherence of $\mathbf{B}_\beta$ converges to the coherence of a conditioned version of $\mathbf{A}$, for $\beta \to 0$.

Lemma B.2. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$*, for* $m \leqslant n$*, have full rank. We have* $\operatorname{coh}(\mathbf{B}_\beta) \to \operatorname{coh}((\mathbf{A}\mathbf{A}^\top)^{-\frac{1}{2}}\mathbf{A})$*, for* $\beta \to 0$*.*

Proof. Define $\mathbf{Q}_\beta = \mathbf{Id}_m + \frac{\mathbf{A}\mathbf{A}^\top}{\beta}$, so that $\mathbf{B}_\beta = \mathbf{Q}_\beta^{-1/2}\mathbf{A}$, and let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the SVD of $\mathbf{A}$. Define $\mathbf{C} = \sqrt{\beta}(\mathbf{A}\mathbf{A}^\top)^{-\frac{1}{2}}\mathbf{A}$ with columns $\mathbf{c}_i$. First note, that

$$\left\|\mathbf{Q}_\beta^{-1} - \left(\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1}\right\| = \max_{i\in[m]}\left|\frac{1}{1+\frac{\sigma_i^2}{\beta}} - \frac{1}{\frac{\sigma_i^2}{\beta}}\right| = \left|\frac{\beta}{\sigma_{\min}^2\left(1+\frac{\sigma_{\min}^2}{\beta}\right)}\right| \leqslant \frac{\beta^2}{\sigma_{\min}^4}$$

and

$$\begin{aligned}\left\|\mathbf{Q}_\beta^{-\frac{1}{2}} - \left(\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-\frac{1}{2}}\right\| &= \max_{i\in[m]}\left|\frac{1}{\sqrt{1+\frac{\sigma_i^2}{\beta}}} - \frac{1}{\sqrt{\frac{\sigma_i^2}{\beta}}}\right| \\ &= \sqrt{\beta}\left|\frac{\sqrt{\sigma_i^2}-\sqrt{\beta+\sigma_i^2}}{\sigma_{\min}\sqrt{\beta+\sigma_{\min}^2}}\right| \leqslant \frac{\beta}{\sigma_{\min}^2}.\end{aligned}$$

Consequently,

$$|\langle \mathbf{b}_i, \mathbf{b}_j\rangle - \langle \mathbf{c}_i, \mathbf{c}_j\rangle| = \left|\mathbf{e}_i^\top \mathbf{A}^\top\left(\mathbf{Q}_\beta^{-1} - \left(\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-1}\right)\mathbf{A}\mathbf{e}_j\right| \leqslant \beta^2\frac{\|\mathbf{A}\|^2}{\sigma_{\min}^4}$$

and

$$|\|\mathbf{b}_i\|_2 - \|\mathbf{c}_i\|_2| \leqslant \|\mathbf{b}_i - \mathbf{c}_i\|_2 = \left\|\left(\mathbf{Q}_\beta^{-\frac{1}{2}} - \left(\frac{\mathbf{A}\mathbf{A}^\top}{\beta}\right)^{-\frac{1}{2}}\right)\mathbf{A}\mathbf{e}_i\right\| \leqslant \beta\frac{\|\mathbf{A}\|}{\sigma_{\min}^2}.$$

Since we have in addition that $\|\mathbf{c}_i\|_2 \leqslant \sqrt{\beta}\|(\mathbf{A}\mathbf{A}^\top)^{-\frac{1}{2}}\mathbf{A}\| = \sqrt{\beta}$, $\|\mathbf{b}_i\|_2 \leqslant \|\mathbf{Q}_\beta^{-\frac{1}{2}}\mathbf{A}\| \leqslant \sqrt{\beta}$, and $\|\mathbf{b}_i\|_2 \geqslant (\|\mathbf{A}\|^2+\beta)^{-\frac{1}{2}}\|\mathbf{a}_i\|_2\sqrt{\beta}$, we get

$$\begin{aligned}&\left|\frac{\langle \mathbf{b}_i, \mathbf{b}_j\rangle}{\|\mathbf{b}_i\|_2\|\mathbf{b}_j\|_2} - \frac{\langle \mathbf{c}_i, \mathbf{c}_j\rangle}{\|\mathbf{c}_i\|_2\|\mathbf{c}_j\|_2}\right| \\ &\quad= \left|\frac{\left(\langle \mathbf{b}_i, \mathbf{b}_j\rangle - \langle \mathbf{c}_i, \mathbf{c}_j\rangle\right)\|\mathbf{c}_i\|_2\|\mathbf{c}_j\|_2 + \langle \mathbf{c}_i, \mathbf{c}_j\rangle\left(\|\mathbf{c}_i\|_2\|\mathbf{c}_j\|_2 - \|\mathbf{b}_i\|_2\|\mathbf{b}_j\|_2\right)}{\|\mathbf{b}_i\|_2\|\mathbf{b}_j\|_2\|\mathbf{c}_i\|_2\|\mathbf{c}_j\|_2}\right|\end{aligned}$$

$$\leqslant \frac{|\langle \mathbf{b}_i, \mathbf{b}_j\rangle - \langle \mathbf{c}_i, \mathbf{c}_j\rangle|}{\|\mathbf{b}_i\|_2\|\mathbf{b}_j\|_2} + \frac{\|\mathbf{c}_i\|_2\,|\|\mathbf{c}_j\|_2 - \|\mathbf{b}_j\|_2| + |\|\mathbf{c}_i\|_2 - \|\mathbf{b}_i\|_2|\,\|\mathbf{b}_j\|_2}{\|\mathbf{b}_i\|_2\|\mathbf{b}_j\|_2}$$

$$= \mathcal{O}(\beta) + \mathcal{O}(\sqrt{\beta}).$$

We conclude by noting that $\mathrm{coh}(\mathbf{C}) = \mathrm{coh}((\mathbf{A}\mathbf{A}^\top)^{-\frac{1}{2}}\mathbf{A})$. □

ORCID iDs

Željko Kereta https://orcid.org/0000-0003-2805-0037

References

[1] Aeron S, Saligrama V and Zhao M 2010 Information theoretic bounds for compressed sensing *IEEE Trans. Inf. Theor.* **56** 5111–30

[2] Arias-Castro E and Eldar Y C 2011 Noise folding in compressed sensing *IEEE Signal Process. Lett.* **18** 478–81

[3] Artina M, Fornasier M and Peter S 2015 Damping noise-folding and enhanced support recovery in compressed sensing *IEEE Trans. Signal Process.* **63** 5990–6002

[4] Attouch H, Bolte J, Redont P and Soubeyran A 2010 Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality *Math. Oper. Res.* **35** 438–57

[5] Attouch H, Bolte J and Svaiter B F 2013 Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods *Math. Program.* **137** 91–129

[6] Bauschke H H and Combettes P L 2011 *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* vol 408 (Berlin: Springer)

[7] Beck A 2017 *First-order Methods in Optimization* (Philadelphia, PA: SIAM)

[8] Bolte J, Nguyen T P, Peypouquet J and Suter B W 2017 From error bounds to the complexity of first-order descent methods for convex functions *Math. Program.* **165** 471–507

[9] Bredies K, Lorenz D A and Reiterer S 2015 Minimization of non-smooth, non-convex functionals by iterative thresholding *J. Optim. Theor. Appl.* **165** 78–112

[10] Candés E J, Wakin M B and Boyd S P 2008 Enhancing sparsity by reweighted l_1 minimization *J. Fourier Anal. Appl.* **14** 877–905

[11] Cormen T H, Leiserson C E, Rivest R L and Stein C 2009 *Introduction to Algorithms* (Cambridge, MA: MIT press)

[12] Daubechies I, Defrise M and Mol C D 2016 Sparsity-enforcing regularisation and ISTA revisited *Inverse Problems* **32** 104001

[13] Fan J and Li R 2001 Variable selection via nonconcave penalized likelihood and its oracle properties *J. Am. Stat. Assoc.* **96** 1348–60

[14] Foucart S and Rauhut H 2013 *A Mathematical Introduction to Compressive Sensing* (Basel: Birkhäuser)

[15] Grasmair M, Klock T and Naumova V 2020 Adaptive multi-penalty regularization based on a generalized lasso path *Appl. Comput. Harmon. Anal.* **49** 30–55

[16] Grasmair M and Naumova V 2016 Conditions on optimal support recovery in unmixing problems by means of multi-penalty regularization *Inverse Problems* **32** 104007

[17] Laude E, Wu T and Cremers D 2018 A nonconvex proximal splitting algorithm under Moreau–Yosida regularization *Int. Conf. on Artificial Intelligence and Statistics (AISTATS)* pp 491–9

[18] Laude E, Wu T and Cremers D 2019 Optimization of inf-convolution regularized nonconvex composite problems *Int. Conf. on Artificial Intelligence and Statistics (AISTATS)* pp 547–56

[19] Li G 2013 Global error bounds for piecewise convex polynomials *Math. Program.* **137** 37–64

[20] Mordukhovich B S 2006 *Variational Analysis and Generalized Differentiation I* (*Basic Theory*) vol 330 (Berlin: Springer)

[21] Naumova V and Peter S 2014 Minimization of multi-penalty functionals by alternating iterative thresholding and optimal parameter choices *Inverse Problems* **30** 125003
[22] Parikh N and Boyd S 2014 Proximal algorithms *Foundations and Trends in Optimization* **1** 127–239
[23] Rockafellar R T and Wets R J-B 2009 *Variational Analysis* vol 317 (Berlin: Springer)
[24] Wang Y, Zeng J, Peng Z, Chang X and Xu Z 2015 Linear convergence of adaptively iterative thresholding algorithms for compressed sensing *IEEE Trans. Signal Process.* **63** 2957–71
[25] Wen F, Chu L, Liu P and Qiu R C 2018 A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning *IEEE Access* **6** 69883–906
[26] Xu Z, Chang X, Xu F and Zhang H 2012 $L_{1/2}$ regularization: a thresholding representation theory and a fast solver *IEEE Trans. Neural Netw. Learn. Syst.* **23** 1013–27
[27] Yu P, Li G and Pong T K 2019 Deducing Kurdyka–Lojasiewicz exponent via inf-projection (arXiv:1902.03635)
[28] Zeng J, Lin S and Xu Z 2016 Sparse regularization: convergence of iterative jumping thresholding algorithm *IEEE Trans. Signal Process.* **64** 5106–18
[29] Zhang C-H 2010 Nearly unbiased variable selection under minimax concave penalty *Ann. Stat.* **38** 894–942