Numer Algor (2018) 78:739–757
DOI 10.1007/s11075-017-0398-6

ORIGINAL PAPER

l_1-l_2 regularization of split feasibility problems

Abdellatif Moudafi[1,2] · Aviv Gibali[3]

Received: 13 March 2017 / Accepted: 10 August 2017 / Published online: 18 August 2017
© Springer Science+Business Media, LLC 2017

Abstract Numerous problems in signal processing and imaging, statistical learning and data mining, or computer vision can be formulated as optimization problems which consist in minimizing a sum of convex functions, not necessarily differentiable, possibly composed with linear operators and that in turn can be transformed to split feasibility problems (SFP); see for example Censor and Elfving (Numer. Algorithms **8**, 221–239 1994). Each function is typically either a data fidelity term or a regularization term enforcing some properties on the solution; see for example Chaux et al. (SIAM J. Imag. Sci. **2**, 730–762 2009) and references therein. In this paper, we are interested in split feasibility problems which can be seen as a general form of Q-Lasso introduced in Alghamdi et al. (2013) that extended the well-known Lasso of Tibshirani (J. R. Stat. Soc. Ser. B **58**, 267–288 1996). Q is a closed convex subset of a Euclidean m-space, for some integer $m \geq 1$, that can be interpreted as the set of errors within given tolerance level when linear measurements are taken to recover a signal/image via the Lasso. Inspired by recent works by Lou and Yan (2016), Xu (IEEE Trans. Neural Netw. Learn. Syst. **23**, 1013–1027 2012), we are interested in a nonconvex regularization of SFP and propose three split algorithms for solving this general case. The first one is based on the DC (difference of convex)

✉ Aviv Gibali
avivg@braude.ac.il

Abdellatif Moudafi
abdellatif.moudafi@univ-amu.fr

1 CNRS-L.I.S UMR, Aix-Marseille Université, 13284, Marseille, France

2 Domaine Universitaire de Saint-Jérôme, Avenue Escadrille Normandie-Niemen, 13397 Marseille, France

3 Department of Mathematics, ORT Braude College, 2161002, Karmiel Israel

algorithm (DCA) introduced by Pham Dinh Tao, the second one is nothing else than the celebrate forward-backward algorithm, and the third one uses a method introduced by Mine and Fukushima. It is worth mentioning that the SFP model a number of applied problems arising from signal/image processing and specially optimization problems for intensity-modulated radiation therapy (IMRT) treatment planning; see for example Censor et al. (Phys. Med. Biol. **51**, 2353–2365, 2006).

Keywords Q-Lasso · Split feasibility · Soft-thresholding · Regularization · DCA algorithm · Forward-backward iterations · Mine-Fukushima algorithm · Douglas-Rachford algorithm

Mathematics Subject Classification (2010) Primary 49J53, 65K10 · Secondary 49M37, 90C25

1 Introduction and preliminaries

Recent developments in science and technology have caused a revolution in data processing, as large datasets are becoming increasingly available and important. To meet the need in big data area, the field of compressive sensing (CS) [13] is rapidly blooming. The process of CS consists of encoding and decoding. The process of encoding involves taking a set of (linear) measurements, $b = Ax$, where A is a matrix of size $m \times n$. If $m < n$, we say the signal $x \in \mathbb{R}^n$ can be compressed. The process of decoding is to recover x from b with an additional assumption that x is sparse. It can be expressed as an optimization problem,

$$\min \|x\|_0 \text{ subject to } Ax = b, \tag{1.1}$$

with $\|\cdot\|_0$ being the l_0 norm, which counts the number of nonzero entries of x; that is

$$\|x\|_0 = |\{x_i \mid x_i \neq 0\}| \tag{1.2}$$

where $|\cdot|$ denotes here the cardinality, i.e., the number of elements of a set. So minimizing the l_0 norm is equivalent to finding the sparsest solution. One of the biggest obstacles in CS is solving the decoding problem above, as l_0 minimization is NP-hard. A popular approach is to replace l_0 by the convex norm l_1, which often gives a satisfactory sparse solution. This l_1 heuristic has been applied in many different fields such as geology and geophysics, spectroscopy, and ultrasound imaging.

Recently, there has been an increase in applying nonconvex metrics as alternative approaches to l_1 . In particular, the nonconvex metric l_p for $p \in (0, 1)$ in [7] can be regarded as a continuation strategy to approximate l_0 as $p \to 0$. The optimization strategies include iterative reweighting [7] and half thresholding [24], and the scale-invariant l_1, formulated as the ratio of l_1 and l_2, was discussed in [14]. Other nonconvex l_1 variants include transformed l_1, sorted l_1, and capped l_1. It is demonstrated in a series of papers [15, 24] that difference of the l_1 and l_2 norms, denoted as l_1-l_2 , outperforms l_1 and l_p in terms of promoting sparsity when sensing matrix A is highly coherent. Based on this idea, we propose the same type of regularization for SFP and propose three splitting algorithms: the first one is nothing but the DC (difference of convex) algorithm (DCA) introduced by Pham Dinh Tao; see for example [19]. The second one is nothing else than the celebrate forward-backward algorithm and the third one uses a method introduced by Mine and Fukushima in [17] for minimizing the sum of a convex function and a differentiable one.

First, remember that the lasso of Tibshirani [22] is the minimization problem

$$\min_{x\in \mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2+\gamma\|x\|_1, \tag{1.3}$$

where A is an $m\times n$ real matrix, $b\in \mathbb{R}^m$, and $\gamma>0$ is a tuning parameter. It is equivalent to the basic pursuit (BP) of Chen et al. [10]

$$\min_{x\in \mathbb{R}^n} \|x\|_1 \quad \text{subject to } Ax=b. \tag{1.4}$$

However, due to errors of measurements, the constraint $Ax=b$ is actually inexact. It turns out that problem (1.4) is reformulated as

$$\min_{x\in \mathbb{R}^n} \|x\|_1 \quad \text{subject to } \|Ax-b\|_p\le\varepsilon, \tag{1.5}$$

where $\varepsilon>0$ is the tolerance level of errors and p is often $1, 2$, or ∞. It is noticed in [1] that if we let $Q:=B_\varepsilon(b)$, the closed ball in $\mathbb{R}^n$ with center b and radius ε, then (1.5) is rewritten as

$$\min_{x\in \mathbb{R}^n} \|x\|_1 \quad \text{subject to } Ax\in Q. \tag{1.6}$$

With Q a nonempty closed convex set of $\mathbb{R}^m$ and P_Q the projection from $\mathbb{R}^m$ onto Q and since that the constraint is equivalent to the condition $Ax-P_Q(Ax)=0$, this leads to the following equivalent Lagrangian formulation

$$\min_{x\in \mathbb{R}^n} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma\|x\|_1, \tag{1.7}$$

with $\gamma>0$ a Lagrangian multiplier. A connection is also made in [1] with the so-called split feasibility problem [5] which is stated as finding x verifying

$$x\in C, \quad Ax\in Q, \tag{1.8}$$

where C and Q are closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. An equivalent minimization formulation of (1.8) is

$$\min_{x\in C} \frac{1}{2}\|(I-P_Q)Ax\|_2^2. \tag{1.9}$$

Its l_1 regularization is given as

$$\min_{x\in C} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma\|x\|_1, \tag{1.10}$$

where $\gamma>0$ is a regularization parameter.

This convex relaxation attracts considerable attention; see for example [1] and references therein. In this paper, we study a nonconvex but Lipschitz continuous metric l_1-l_2 for SFP. As illustrated in [15], the level curves of l_1-l_2 are closer to l_0 than those of l_1, which motivated us to consider the nonconvex l_1-l_2 regularization for split feasibility problem, namely

$$\min_{x\in C} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma(\|x\|_1-\|x\|_2), \tag{1.11}$$

and propose three algorithms. The first uses the DCA which is a descent method without line search introduced by Tao and An [19] for minimizing a function f which is the difference of two lower semicontinuous proper convex functions g and h on the space $I\!R^n$. The second one is based on the gradient proximal method to solve the problem (1.11) by full splitting, that is, at every iteration, the only operations involved are evaluations of the gradient of the function $\frac{1}{2}\|(I - P_Q)A(\cdot)\|_2^2$, the proximal mapping of $\|\cdot\|_1 - \|\cdot\|_2$, A, or its transpose A^t. The third one is based on an algorithm for minimizing the sum of a convex function and a differentiable one introduced by Mine and Fukushima in [17].

In [1], properties and iterative methods for (1.7) are investigated. Remember also that many authors devoted their works to the unconstrained minimization problem $\min_{x\in H} f_1(x) + f_2(x)$ with f_1, f_2 are two proper, convex lower semi continuous functions defined on a Hilbert space H and f_2 differentiable with a β-Lipschitz continuous gradient for some $\beta > 0$ and an effective method to solve it is the forward-backward algorithm which from an initial value x_0 generates a sequence (x_k) by the following iteration

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k prox_{\gamma_k f_1}(x_k - \gamma_k \nabla f_2(x_k)), \tag{1.12}$$

where $\gamma_k > 0$ is the algorithm step size, $0 < \lambda_k < 1$ is a relaxation parameter, and $prox_{\gamma_k f_1}$ being the proximal mapping defined in (2.40).

It is well-known, see for instance [11], that if (γ_k) is bounded and (λ_k) is bounded from below, then (x_k) weakly converges to a solution of $\min_{x\in H} f_1(x) + f_2(x)$ provided that the set of solutions is nonempty.

In order to relax the assumption on the differentiability of f_2, the Douglas-Rachford algorithm was introduced. It generates a sequence (y_k) as follows:

$$\begin{cases} y_{k+1/2} = prox_{\kappa f_2} y_k; \\ y_{k+1} = y_k + \tau_k\big(prox_{\kappa f_1}(2y_{k+1/2} - y_k) - y_{k+1/2}\big) \end{cases} \tag{1.13}$$

where $\kappa > 0$, (τ_k) is a sequence of positive reals. It is well-known that (y_k) converges weakly to y such that $prox_{\kappa f_2} y$ is a solution of the unconstrained minimization problem above provided that $\forall k \in I\!N,\ \tau_k \in]0, 2[$ and $\sum_{k=0}^{\infty} \tau_k(2 - \tau_k) = +\infty$ and the set of solutions is nonempty.

In what follows, we are interested in (1.11) which is more challenging and we will focus our attention on the algorithmic aspect.

Our paper is organized as follows. In Section 2, we first start with definitions and notions which are needed for the presentation of our three proposed schemes, the DCA algorithm, the forward-backward algorithm, and the third based on the Mine-Fukushima algorithm. We also give full convergence theorem for the proposed schemes. Later in Section 3, we present several numerical experiments which illustrates the performances of our schemes compared with the CQ and relaxed CQ algorithms. We include random linear system of equations as well as an example in sparse signal recovery. Finally, in Section 4, we provide further insights into how to compute the proximal mapping of a sum of two functions by coupling the Douglas-Rachford and the forward-backward algorithms.

2 Computational approaches

2.1 DCA

First, remember that the *subdifferential set* (or just subdifferential) of a convex function h is defined as

$$\partial h(x) := \{u \in I\!R^n; h(y) \geq h(x) + \langle u, y - x\rangle \ \forall y \in I\!R^n\}. \tag{2.1}$$

Each element of $\partial h(x)$ is called *subgradient*. In case that the function h is continuously differentiable then $\partial h(x) = \{\nabla h(x)\}$, this is the *gradient* of h. It is easily seen that

$$\partial \frac{1}{2}\|Ax - y\|^2 = \nabla \frac{1}{2}\|Ax - y\|^2 = A^t(Ax - y), \tag{2.2}$$

and

$$(\partial \|x\|_1)_i = \begin{cases} sgn(x_i) & \text{if } x_i \neq 0; \\ \text{any element of } [-1, 1] & \text{if } x_i = 0. \end{cases} \tag{2.3}$$

The *characteristic function* of a set $C \subseteq I\!R^n$ is defined as

$$i_C(x) = \begin{cases} 0 & \text{if } x \in C; \\ +\infty & \text{otherwise} \end{cases} \tag{2.4}$$

such function is convenient to enforce hard constraints on the solution. Moreover, the *normal cone* of C at $x \in C$, denoted by $N_C\,(x)$, is defined

$$N_C\,(x) := \{d \in I\!R^n \mid \langle d, y - x\rangle \leq 0, \ \forall y \in C\}. \tag{2.5}$$

A known relation between the above definition is that $\partial i_C = N_C$. Another useful definition which will be useful in the sequel is the following. A sequence (x_k) is called *asymptotically regular*, if $\lim_{n\to\infty} \|x_{k+1} - x_k\| = 0$.

For finding critical points of $f := g - h$, the DCA involves the construction of two sequences (x_k) and (y_k) by the following rules

$$\begin{cases} y_k \in \partial h(x_k); \\ x_{k+1} = argmin_{x \in I\!R^n}\big(g(x) - (h(x_k) + \langle y_k, x - x_k\rangle)\big). \end{cases} \tag{2.6}$$

Note that by the definition of subdifferential , we can write

$$h(x_{k+1}) \geq h(x_k) + \langle y_k, x_{k+1} - x_k\rangle. \tag{2.7}$$

Since x_{k+1} minimizes $g(x) - (h(x_k) + \langle y_k, x - x_k\rangle)$, we also have

$$g(x_{k+1}) - (h(x_k) + \langle y_k, x_{k+1} - x_k\rangle) \leq g(x_k) - h(x_k). \tag{2.8}$$

Combining the last inequalities, we obtain

$$f(x_k) = g(x_k) - h(x_k) \geq g(x_{k+1}) - (h(x_k) + \langle y_k, x_{k+1} - x_k\rangle) \geq f(x_{k+1}). \tag{2.9}$$

Therefore, the DCA provides a monotonically decreasing sequence $(f(x_k))$ which converges provided that the objective function f is bounded below.

The objective function in (1.11) has the following DC decomposition

$$\min_{x \in I\!R^n} \left(\frac{1}{2}\|(I - P_Q)Ax\|_2^2 + \gamma\|x\|_1 + i_C(x)\right) - \gamma\|x\|_2. \tag{2.10}$$

Observe that $\|x\|_2$ is differentiable with gradient $x/\|x\|_2$ for any $x \neq 0$ and we also have $0 \in \partial\|\cdot\|_2(0)$, which leads to the following iterates

$$x_{k+1}=\begin{cases} argmin_{x\in I\!R^n} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma\|x\|_1+i_C(x) & \text{if } x_k=0\\ argmin_{x\in I\!R^n} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma\|x\|_1+i_C(x)-\left\langle x,\gamma\frac{x_k}{\|x_k\|_2}\right\rangle & \text{if } x_k\neq 0, \end{cases} \tag{2.11}$$

obtained by setting in the rules (2.6): $g(x) = 1/2\|(I-P_Q)Ax\|_2^2 + \gamma\|x\|_1 + i_C(x)$ and $h(x) = \gamma\|x\|_2$. (2.11) is equivalent, using the definition of the characteristic function, to

$$x_{k+1}=\begin{cases} argmin_{x\in C} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma\|x\|_1 & \text{if } x_k=0\\ argmin_{x\in C} \frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma\|x\|_1-\langle x,\gamma\frac{x_k}{\|x_k\|_2}\rangle & \text{if } x_k\neq 0. \end{cases} \tag{2.12}$$

Now, we define for all $\gamma > 0$, the following function

$$\Gamma(x)=\frac{1}{2}\|(I-P_Q)Ax\|_2^2+\gamma(\|x\|_1-\|x\|_2)+i_C(x). \tag{2.13}$$

We are in a position to prove the following convergence properties of the iterative step (2.11):

Proposition 2.1 *Let (x_k) be the sequence generated by Algorithm 2.11.*

(i) *For all $\gamma > 0$ we have that $\lim_{\|x\|_2\to+\infty}\Gamma(x) = +\infty$. Γ is therefore coercive in the sense that its level sets are bounded, namely $\{x \in I\!R^n; \Gamma(x) \le \Gamma(x_0)\}$ is bounded for any $x_0 \in I\!R^n$.*

(ii) *The sequence (x_k) is bounded.*

(iii) *If $\lim_{k\to+\infty}\|x_{k+1}-x_k\|_2 = 0$, i.e., (x_k) is asymptotically regular, then any nonzero limit point x^* of the sequence (x_k) is a stationary point of (1.11), namely*

$$0 \in A^t(I-P_Q)Ax^*+\gamma\left(\partial\|x^*\|_1-\frac{x^*}{\|x^*\|_2}\right)+N_C(x^*). \tag{2.14}$$

Proof Recall first that the support of x is defined by $\text{supp}(x) = \{1 \le i \le n; x_i \neq 0\}$ and that $\|x\|_0 = |\text{supp}(x)|$ is the cardinality of $\text{supp}(x)$. To prove (i)–(ii), remember that for all $x \neq 0$, we have $\|x\|_1 - \|x\|_2 \ge 0$ and that $\|x\|_1 - \|x\|_2 = 0 \Leftrightarrow \|x\|_0 = 1$. With this fact in hand, we can easily verify that Γ is coercive.

Now, a simple computation which uses the fact that $\|a\|^2 - \|b\|^2 = \|a-b\|^2 + 2\langle b, a-b\rangle$, gives

$$\begin{aligned}\Gamma(x_k)-\Gamma(x_{k+1}) = {} & \frac{1}{2}\|Ax_k-Ax_{k+1}-\big(P_Q(Ax_k)-P_Q(Ax_{k+1})\big)\|^2\\ & +\langle Ax_k-Ax_{k+1}-\big(P_Q(Ax_k)-P_Q(Ax_{k+1})\big), Ax_{k+1}\\ & \quad -P_Q(Ax_{k+1})\rangle\\ & +\gamma(\|x_k\|_1-\|x_{k+1}\|_1-\|x_k\|_2+\|x_{k+1}\|_2). \end{aligned} \tag{2.15}$$

The first-order optimality condition at x_{k+1} as the solution of the problem (2.11) and the fact that $\partial(\|\cdot\|_1 + i_C)(x) = \partial\|x\|_1 + N_C(x)$ (since a norm is continuous) lead to.

$$A^t(I - P_Q)Ax_{k+1} + \gamma(w_{k+1} - y_k) + p_{k+1} = 0,$$

with $y_k \in \partial\|x_k\|_2$, $w_{k+1} \in \partial\|x_{k+1}\|_1$ and $p_{k+1} \in N_C(x_{k+1})$. This combined with $\langle w_k, x_{k+1}\rangle = \|x_{k+1}\|_1$ gives

$$\langle A(x_k - x_{k+1}), (I - P_Q)Ax_{k+1}\rangle + \gamma(\langle w_{k+1}, x_k\rangle - \|x_{k+1}\|_1 + \langle y_k, x_{k+1} - x_k\rangle) \\ -\langle p_{k+1}, x_{k+1} - x_k\rangle = 0. \tag{2.16}$$

Combining (2.15) and (2.16), we can write

$$\begin{aligned}\Gamma(x_k) - \Gamma(x_{k+1}) = {} & \frac{1}{2}\|Ax_k - Ax_{k+1} - \big(P_Q(Ax_k) - P_Q(Ax_{k+1})\big)\|^2 \\ & -\gamma\big(\langle w_{k+1}, x_k\rangle - \|x_{k+1}\|_1 + \langle y_k, x_{k+1} - x_k\rangle\big) \\ & +\langle p_{k+1}, x_{k+1} - x_k\rangle \\ & -\langle Ax_{k+1} - P_Q(Ax_{k+1}), P_Q(Ax_k) - P_Q(Ax_{k+1})\rangle \\ & +\gamma(\|x_k\|_1 - \|x_{k+1}\|_1 - \|x_k\|_2 + \|x_{k+1}\|_2)\rangle.\end{aligned} \tag{2.17}$$

The characterization of the orthogonal projection, namely

$$\langle x - P_Q(x), z - P_Q(x)\rangle \le 0 \; \forall z \in Q, \tag{2.18}$$

assures that

$$\langle (I - P_Q)Ax_{k+1}, P_Q(Ax_k) - P_Q(Ax_{k+1})\rangle \le 0, \tag{2.19}$$

and thus

$$\Gamma(x_k) - \Gamma(x_{k+1}) \ge \frac{1}{2}\|(I - P_Q)(Ax_k)(I - P_Q)(Ax_{k+1})\|^2 + \gamma(\|x_k\|_1 - \langle w_{k+1}, x_k\rangle) \\ +\gamma(\|x_{k+1}\|_2 - \|x_k\|_2 - \langle y_k, x_{k+1} - x_k\rangle) + \langle p_{k+1}, x_{k+1} - x_k\rangle.$$

On the other hand, since $|w_{k+1,i}| \le 1$ for $i = 1, ..., n$, $y_k \in \partial\|x_k\|_2$ and $p_{k+1} \in N_C(x_{k+1})$, we also have

$$\|x_k\|_1 - \langle w_{k+1}, x_k\rangle \ge 0 \; \|x_{k+1}\|_2 - \|x_k\|_2 - \langle y_k, x_{k+1} - x_k\rangle \ge 0 \text{ and } \langle p_{k+1}, x_{k+1} - x_k\rangle \ge 0. \tag{2.20}$$

Consequently,

$$\Gamma(x_k) - \Gamma(x_{k+1}) \ge \frac{1}{2}\|(I - P_Q)(Ax_k) - (I - P_Q)(Ax_{k+1})\|^2 \\ +\gamma(\|x_{k+1}\|_2 - \|x_k\|_2 - \langle y_k, x_{k+1} - x_k\rangle) \ge 0. \tag{2.21}$$

This ensures that the sequence $(\Gamma(x_k))$ is monotonically decreasing, which in turn ensures that the sequence $(x_k) \subset \{x \in I\!R^n, \Gamma(x) \le \Gamma(x_0)\}$ that is bounded since Γ is coercive.

(iii) If $x_1 = x_0 = 0$, we then stop the algorithm producing the solution $x^* = 0$. Otherwise, it follows from (2.21)

$$\Gamma(x_0) - \Gamma(x_1) \ge \gamma\|x_1\|_2 > 0, \tag{2.22}$$

so $x_k \neq 0$ for all $k \geq 1$. Since $(\Gamma(x_k))$ is convergent, substituting $y_k = \frac{x_k}{\|x_k\|_2}$ leads to

$$\lim_{k\to+\infty} \|(I-P_Q)(Ax_k)-(I-P_Q)(Ax_{k+1})\|^2=0 \text{ and } \lim_{k\to+\infty} \|x_{k+1}\|_2-\frac{\langle x_k, x_{k+1}\rangle}{\|x_k\|_2}=0. \tag{2.23}$$

Now, let (x_{k_ν}) be a subsequence of (x_k) converging to $x^* \neq 0$, so the optimality condition at the k_ν the step of Algorithm (2.11) reads

$$-\left(A^t(I-P_Q)Ax_{k_\nu} - \gamma\frac{x_{k_\nu-1}}{\|x_{k_\nu-1}\|_2}\right) \in \gamma\partial\|x_{k_\nu}\|_1 + N_C(x_{k_\nu}). \tag{2.24}$$

Since $\lim_{\nu\to+\infty} x_{k_\nu} = x^*$, the operator $A^t(I-P_Q)A$ is Lipschitz continuous, the sequence (x_k) is assumed to be asymptotically regular, and x^* is away from 0, we have

$$\begin{aligned}\lim_{\nu\to+\infty}\left(A^t(I-P_Q)Ax_{k_\nu} - \gamma\frac{x_{k_\nu-1}}{\|x_{k_\nu-1}\|_2}\right) &= \lim_{\nu\to+\infty}\left(A^t(I-P_Q)Ax_{k_\nu} - \gamma\frac{x_{k_\nu}}{\|x_{k_\nu}\|_2}\right.\\ &\quad\left.+\gamma\left(\frac{x_{k_\nu}}{\|x_{k_\nu}\|_2}-\frac{x_{k_\nu-1}}{\|x_{k_\nu-1}\|_2}\right)\right)\\ &= A^t(I-P_Q)Ax^* - \gamma\frac{x^*}{\|x^*\|_2},\end{aligned} \tag{2.25}$$

and by passing to the limit as $\nu \to +\infty$ in (2.24) and by taking into account the fact that $\partial(\|\cdot\|_1 + i_C)$ is a maximal monotone operator, which assures that its graph is closed, we obtain at the limit

$$-\left(A^t(I-P_Q)Ax^* - \gamma\frac{x^*}{\|x^*\|_2}\right) \in \gamma\partial\|x^*\|_1 + N_C(x^*), \tag{2.26}$$

in other words, x^* is a stationary point. □

The asymptotical regularity assumption is satisfied in the particular case where Q is a singleton considered in [23]. In what follows, we will prove that it is also the case in the interesting setting of closed convex cones which usually arises, for example, in statistical applications and also in image recovery where subspaces are often used. Likewise, when the projection has the nice property to be homogeneous with respect to the set Q, which is the case, for instance, for balls, rectangles,... when the points to project are outside.

Proposition 2.2 *The iteration sequence is asymptotically regular in the following three cases:*

i) $Q = \{b\}$.
ii) *Q is a closed convex cone and when Q is a subspace.*
iii) *The projection is a non-negative homogeneous function with respect to the set Q, namely*

$$\forall\alpha > 0\ \forall x \in I\!R^n \text{ one has } P_{\alpha Q}(x) = \alpha P_Q(x). \tag{2.27}$$

Proof i) Indeed, in this case, relation (2.21) reduces to

$$\Gamma(x_k) - \Gamma(x_{k+1}) \geq \frac{1}{2}\|Ax_k - Ax_{k+1}\|^2 \\ +\gamma(\|x_{k+1}\|_2 - \|x_k\|_2 - \langle y_k, x_{k+1} - x_k\rangle) \geq 0, \quad (2.28)$$

which is exactly the relation that gives the asymptotical regularity in [23]. Following the same lines of the proof of [23]-Proposition 3.1-(b), we obtain the desired result which is similar to the end of the proof of iii) below.

ii) In this setting, the projection is a non-negative homogeneous function, i.e.,

$$\forall \alpha \geq 0 \; \forall x \in I\!R^n \text{ one has } P_Q(\alpha x) = \alpha P_Q(x). \quad (2.29)$$

At this stage, observe that this property holds true also for subspaces since the projection is linear in this case and the proof will be the same. Now, remember that $I - P_Q = P_{Q^*}$, where $Q^* := \{y \in I\!R^n, \langle y, x\rangle \leq 0 \; \forall x \in Q\}$ and set $c_k = \frac{\langle x_k, x_{k+1}\rangle}{\|x_k\|^2}$ and $\varepsilon_k = x_{k+1} - c_k x_k$. It suffices then to prove that $\lim_{k\to+\infty} \varepsilon_k = 0$ and $\lim_{k\to+\infty} c_k = 1$. A simple computation shows that

$$\|\varepsilon_k\|_2^2 = \|x_{k+1}\|_2^2 - \frac{\langle x_k, x_{k+1}\rangle^2}{\|x_k\|^2} \to 0 \text{ as } k \to +\infty, \quad (2.30)$$

by virtue of the second limit in (2.23). On the other hand, using the first limit in (2.23), we can write

$$\begin{aligned} 0 = \lim_{k\to+\infty} \|P_{Q^*}(Ax_k) - P_{Q^*}(Ax_{k+1})\| &= \lim_{k\to+\infty} \|P_{Q^*}(Ax_k) - P_{Q^*}(A(c_k x_k + \varepsilon_k))\| \\ &= \lim_{k\to+\infty} \|P_{Q^*}(Ax_k) - P_{Q^*}(A(c_k x_k))\| \\ &= \lim_{k\to+\infty} \|P_{Q^*}(Ax_k) - P_{Q^*}(c_k A(x_k))\| \\ &= \lim_{k\to+\infty} |c_k - 1| \|P_{Q^*}(Ax_k)\|, \end{aligned} \quad (2.31)$$

where we used the homogeneity of the projection and the fact that $c_k > 0$. The latter follows from the fact that x_{k+1} is a minimizer in Algorithm 2.11. More precisely, we can write

$$\frac{1}{2}\|P_{Q^*}(Ax_{k+1})\|_2^2 + \gamma\|x_{k+1}\|_1 - \langle x_{k+1}, \gamma\frac{x_k}{\|x_k\|_2}\rangle \leq \frac{1}{2}\|P_{Q^*}(A(0))\|_2^2 \\ +\gamma\|0\|_1 - \langle 0, \gamma\frac{x_k}{\|x_k\|_2}\rangle = 0. \quad (2.32)$$

From which, we obtain that $c_k > 0$. Now, if $\lim_{k\to+\infty}(c_k - 1) \neq 0$, then there exists a subsequence (x_{k_ν}) such that $\lim_{\nu\to+\infty} P_{Q^*}(Ax_{k_\nu}) = 0$. So, we have

$$\lim_{\nu\to+\infty} \Gamma(x_{k_\nu}) \geq \lim_{\nu\to+\infty} \frac{1}{2}\|P_{Q^*}(Ax_{k_\nu})\|^2 = 0 = \Gamma(x_0), \quad (2.33)$$

which contradicts the fact that

$$\Gamma(x_{k_\nu}) \leq \Gamma(x_1) < \Gamma(x_0) \; \forall k_\nu \geq 1. \quad (2.34)$$

Consequently, $\lim_{k\to+\infty} c_k = 1$ and thus $\lim_{k\to+\infty} \|x_{k+1} - x_k\| = 0$ which completes the proof.

iii) To begin with, a simple calculation shows that $P_{\alpha Q}(x) = \alpha P_Q(\frac{1}{\alpha}x)$; see for example [6, Lemma 2.1] with $U = I$ and $A = 0$. Hence, we have $P_Q(c_k(Ax_k)) = c_k P_{\frac{1}{c_k}Q}(Ax_k)$ and thus

$$P_Q(c_k(Ax_k)) = c_k P_{\frac{1}{c_k}Q}(Ax_k) = c_k \frac{1}{c_k} P_Q(Ax_k) = P_Q(Ax_k), \tag{2.35}$$

by virtue of the homogeneous property of the projection and the fact that $c_k > 0$. With this and the first limit in (2.23) in hand, we can successively write

$$\begin{aligned} \lim_{k\to+\infty} \|(I-P_Q)(Ax_k) - (I-P_Q)(Ax_{k+1})\| &= \lim_{k\to+\infty} \|(I-P_Q)(Ax_k) \\ &\quad -(I-P_Q)(A(c_kx_k+\varepsilon_k))\| \\ &= \lim_{k\to+\infty} \|(I-P_Q)(Ax_k) \\ &\quad -(I-P_Q)(A(c_kx_k))\| \\ &= \lim_{k\to+\infty} \|(I-P_Q)(Ax_k) \\ &\quad -(I-P_Q)(c_kA(x_k))\| \\ &= \lim_{k\to+\infty} |c_k-1|\|Ax_k\| = 0. \end{aligned} \tag{2.36}$$

Now, if $\lim_{k\to+\infty}(c_k-1) \neq 0$, then there exists a subsequence (x_{k_ν}) of (x_k) such that $\lim_{\nu\to+\infty} Ax_{k_\nu} = 0$. So, we have

$$\lim_{\nu\to+\infty} \Gamma(x_{k_\nu}) \geq \lim_{\nu\to+\infty} \frac{1}{2}\|(I-P_Q)(Ax_{k_\nu})\|^2 = \frac{1}{2}\|P_Q(0)\|^2 = \Gamma(x_0), \tag{2.37}$$

which contradicts the fact that

$$\Gamma(x_{k_\nu}) \leq \Gamma(x_1) < \Gamma(x_0) \ \forall k_\nu \geq 1. \tag{2.38}$$

Consequently, $\lim_{k\to+\infty} c_k = 1$ and again the sequence (x_k) is asymptotically regular. □

Remark 2.1 Each DCA iteration requires solving a l_1-regularized split feasibility subproblem of the form

$$\min_{x\in C} \left(\frac{1}{2}\|(I-P_Q)Ax\|_2^2 + \langle x, v\rangle + \gamma\|x\|_1 \right), \tag{2.39}$$

where $v \in I\!R^n$ is a constant vector. This problem can be done by the two split proximal algorithms (coupling the forward-backward and Douglas-Rachford algorithms) proposed in [18, 23] and also by the alternating direction method of multipliers (ADMM) following the analysis developed in [22] for the special case where Q is a singleton. The details will be given in the Appendix.

2.2 Forward-backward splitting algorithm

To begin with, recall that the proximal mapping (or the Moreau envelope) of a proper, convex, and lower semicontinuous function φ of parameter $\lambda > 0$ is defined by

$$prox_{\lambda\varphi}(x) := arg \min_{v\in I\!R^n} \left\{ \varphi(v) + \frac{1}{2\lambda}\|v - x\|^2 \right\}, \ x \in I\!R^n, \tag{2.40}$$

and that it has closed-form expression in some important cases. For example, if $\varphi = \|\cdot\|_1$, then for $x \in I\!R^n$

$$prox_{\lambda\|\cdot\|_1}(x) = (prox_{\lambda|\cdot|}(x_1), prox_{\lambda|\cdot|}(x_n)), \tag{2.41}$$

where $prox_{\lambda|\cdot|}(x_k) = sgn(x_k)\max_{k=1,2,\cdots n}\{|x_k| - \lambda, 0\}$.

If $\varphi = i_C$, we have

$$prox_{\gamma\varphi}(x) = Proj_C(x) := arg \min_{z\in C}\|x - z\|. \tag{2.42}$$

For sake of simplicity and clarity, we set in what follows $C = I\!R^n$. Observe that when $\gamma > 0$, the minimization problem (1.11) can be written as

$$\min_{x\in I\!R^n} \frac{1}{2\gamma}\|(I - P_Q)Ax\|_2^2 + \|x\|_1 - \|x\|_2. \tag{2.43}$$

It is worth mentioning that when $C \neq I\!R^n$, this requires to compute the proximal operator of a sum, namely $prox_{i_C+\gamma_k(\|\cdot\|_1-\|\cdot\|_2)}$ which may be performed with Douglas-Rachford iterations in the spirit of the analysis developed in [9] and [18].

A closed-form solution of $prox_{\|x\|_1-\|x\|_2}$ was proposed in [15]; in particular, we have the following lemma.

Lemma 2.3 *Given $y \in I\!R^n$, $\lambda > 0$ and setting $r(x) = \|\cdot\|_1 - \|\cdot\|_2$, we have*

(i) *When $\lambda < \|y\|_\infty$, then*

$$prox_{\lambda r}(y) = \frac{\lambda + \|prox_{\lambda\|\cdot\|_1} y\|_2}{\|prox_{\lambda\|\cdot\|_1} y\|_2} prox_{\lambda\|\cdot\|_1} y. \tag{2.44}$$

(ii) *When $\lambda = \|y\|_\infty$, then $x^* \in prox_{\lambda r}(y)$ if and only if it satisfies $x_i^* = 0$ if $|y_i| < \lambda$, $\|x^*\|_2 = \lambda$ and $x_i^* y_i \geq 0$ for all i.*

(iii) *When $\lambda > \|y\|_\infty$, then $x^* \in prox_{\lambda r}(y)$ if and only if it is a 1-spare vector satisfying $x_i^* = 0$ if $|y_i| < \|y\|_\infty$, $\|x^*\|_2 = \|y\|_\infty$ and $x_i^* y_i \geq 0$ for all i.*

By setting $l(x) = \frac{1}{2\gamma}\|(I - P_Q)Ax\|_2^2$, the forward-backward splitting algorithm can be expressed as follows:

$$x_{k+1} \in prox_{\lambda r}(x_k - \lambda\nabla l(x_k)). \tag{2.45}$$

Since the two assumptions of [15, Theorem 3] are satisfied, namely the coerciveness of the objective function and differentiability of the function l with Lipschitz-continuity of its gradient, a direct application of this theorem leads to the following convergence result:

Proposition 2.4 *If $\lambda < \frac{\gamma}{\|A\|^2}$, then the objective values are decreasing and there exists a subsequence of (x_k) that converges to a stationary point. Furthermore, any limit point of (x_k) is a stationary point.*

2.3 Mine-Fukushima algorithm

At this stage, we would like to mention that in the case where C is strictly convex and that we can generate from an initial point x_0 a sequence x_k such that $x_k \neq 0$ for all $k \in I\!N$, then the algorithm introduced by Mine-Fukushima in [17] is applicable. Indeed, problem (1.11) can be written as

$$\min_{x \in I\!R^n} (\phi(x) := f(x) + g(x)), \tag{2.46}$$

with $f(x) = \frac{1}{2}\|(I - P_Q)Ax\|_2^2 - \gamma\|x\|_2$ and $g(x) = \gamma\|x\|_1 + i_C(x)$. Observe that in this case, we have for $x \neq 0$, that $\nabla f(x) = A^t(I - P_Q)Ax - \gamma\frac{x}{\|x\|_2}$ and $\partial g(x) = \partial\|x\|_1 + N_C(x)$.

So [17, Algorithm 2.1] take the following from

Algorithm (Mine-Fukushima):

Step 1. Let x_0 be any initial point. Set $k = 0$, and go to step 2.
Step 2. If $-\nabla f(x_k) \in \partial g(x_k)$, stop. Otherwise, go to step 3.
Step 3. Find a minimum $\tilde{x}_k$ of

$$\min_{x \in C} \left(\left\langle x, A^t(I - P_Q)Ax_k - \gamma\frac{x_k}{\|x_k\|_2} \right\rangle + \gamma\|x\|_1 \right), \tag{2.47}$$

and go to step 4.
Step 4. Find

$$x_{k+1} = \lambda_k\tilde{x}_k + (1 - \lambda_k)x_k, \tag{2.48}$$

such that $\lambda_k \geq 0$ and

$$\phi(x_{k+1}) \leq \phi(\lambda\tilde{x}_k + (1 - \lambda)x_k) \text{ for all } \lambda \geq 0. \tag{2.49}$$

Set $k = k + 1$, and go to step 2.

Observe that solving (2.47) in step 3 is equivalent to finding $\tilde{x}_k$ such that $-\nabla f(x_k) \in \partial g(\tilde{x}_k)$.

Since ϕ is coercive in our case, a direct application of [17, Theorem 3.] yields the following result.

Proposition 2.5 *The sequence (x_k) generated by the Mine-Fukushima Algorithm contains a subsequence which converges to a critical point x^* of (2.46), namely*

$$-A^t(I - P_Q)Ax^* - \gamma\frac{x^*}{\|x^*\|_2} \in \partial\|x^*\|_1 + N_C(x^*). \tag{2.50}$$

Remark 2.2 Theassumption of strict convexity on the convex set C can be removed by applying the following process: for some $\mu > 0$ consider the following decomposition of the objective function ϕ: $\phi = \tilde{f} + \tilde{g}$ with $\tilde{f}(x) = f(x) - \mu\frac{\|x\|_2^2}{2}$ and g by $\tilde{g}(x) = g(x) + \mu\frac{\|x\|_2^2}{2}$. Relation (2.47) becomes

$$\min_{x\in C}\left(\left\langle x, A^t(I-P_Q)Ax_k + \mu x_k - \gamma\frac{x_k}{\|x_k\|_2}\right\rangle + \gamma\|x\|_1 + \mu\frac{\|x\|_2^2}{2}\right). \tag{2.51}$$

3 Numerical experiments

In this section, we present two numerical examples demonstrating the performances of our proposed schemes. In both experiments, we wish to solve the linear system of equations: $Ax = b$ with $A \in I\!R^{120\times 512}$. In the first example, we generate 50 random problems from a normal distribution with mean zero and variance one. For the second experiment, we choose a problem in the field of compressed sensing, which consists of recovering a sparse signal $x \in I\!R^{512}$ with 50 nonzero elements from 120 measurements. In this case, we also include noise, that is, we wish to solve $Ax = b + \varepsilon$, where ε is the noise with bounded variance 10^{-4}.

For the comparison of our proposed schemes, we decided also to include Byrne CQ algorithm [2, 3] and Qu and Xiu [20]-modified CQ algorithm. Byrne CQ algorithm is designed to solve $Ax = b$, and hence, we choose $C = I\!R^n_+$ and $Q = \{b\}$. The CQ iterative step reads as follows:

$$x_{k+1} = P_C(x_k - \widehat{\gamma}A^t(I-P_Q)Ax_k) \tag{3.1}$$

and for the specific choice of C and Q, it translates to

$$x_{k+1} = \left(x_k - \widehat{\gamma}A^t(Ax_k - b)\right)_+ \tag{3.2}$$

and it is denoted in our plots (Figs. 1 and 2) as CQ.

Qu and Xiu [20]-modified CQ algorithm (see also Tang et al. [21]) uses subgradient (elements of the subdifferential set) projection onto super-sets $C \subseteq C_k$ and $Q \subseteq Q_k$ instead of the orthogonal projections onto C and Q . The algorithm also makes use of adaptive step size α_k instead of fixed $\widehat{\gamma}$ as in the CQ algorithm. The algorithm is as follows.

Algorithm (modified CQ):

Step 1. Given constants $l, \mu \in (0, 1)$ and choose $x_0 \in I\!R^n$. Set $k = 0$, and go to step 2.
Step 2. Given the current iterate x_k, let

$$\overline{x_k} = P_{C_k}(x_k - \alpha_k A^t(I-P_Q)Ax_k) \tag{3.3}$$

where $\alpha_k l^{m_k}$ and m_k is the smallest non-negative integer m such that

$$\|A^t(I-P_Q)Ax_k - A^t(I-P_Q)A\overline{x_k}\| \le \mu\frac{\|x_k - \overline{x_k}\|}{\alpha_k}. \tag{3.4}$$

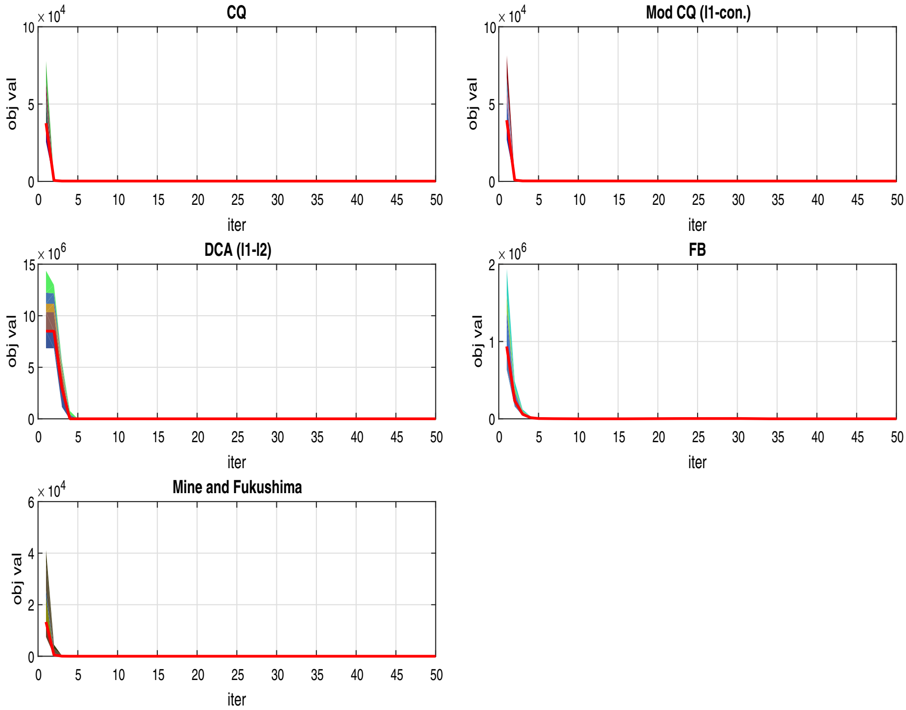

Fig. 1 Testing our proposed algorithms for 50 random problems $Ax = b$, where $A \in I\!R^{120\times 512}$

And the next iterate is calculated via

$$x_{k+1} = P_{C_k}(x_k - \alpha_k A^t (I - P_Q)A\overline{x_k}). \tag{3.5}$$

Set $k = k + 1$, and go to step 2.

While for the CQ algorithm, we wish to solve $Ax = b$, for the modified CQ algorithm, we wish to the consider $Ax = b$ with $l1$ regularization; this is known as the LASSO problem [22] (strongly related to the Basis Pursuit denoising problem [8])

$$\min_{x\in C} \frac{1}{2}\|Ax - b\|_2^2 \quad \text{subject to } \|x\|_1 \le t \tag{3.6}$$

where $t > 0$ is a given constant. So in this case, we choose $C = \{x \mid \|x\|_1 \le t\}$ and $Q = \{b\}$. We define the convex function $c(x) = \|x\|_1 - t$ and denote the level set C_k by

$$C_k = \{x \mid c(x^k) + \langle \xi_k, x - x^k\rangle \le 0\}, \tag{3.7}$$

where $\xi_k \in \partial c(x_k)$ is an element (subgradient) from the subdifferential of c at x_k. The orthogonal projection onto C_k can be calculated by the following,

$$P_{C_k}(y) = \begin{cases} y, & \text{if } c(x_k) + \langle \xi_k, y - x_k\rangle \le 0, \\ y - \frac{c(x_k)+\langle \xi_k, y-x_k\rangle}{\|\xi_k\|^2}\xi_k, & \text{otherwise.} \end{cases} \tag{3.8}$$

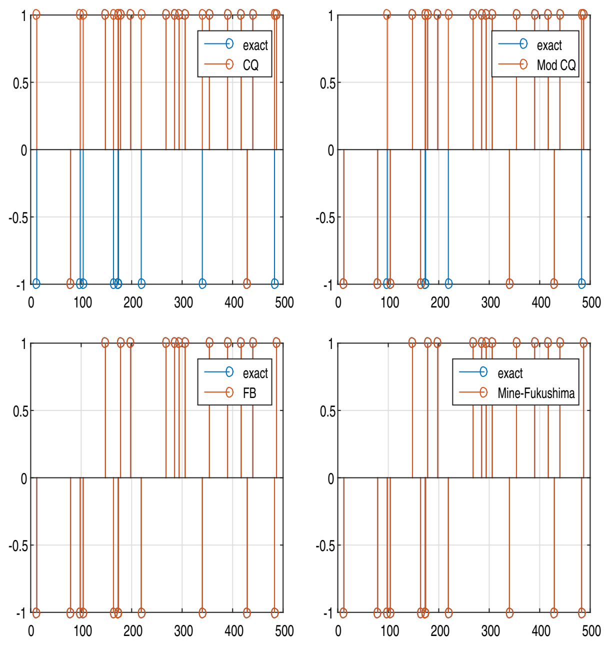

Fig. 2 Testing our proposed algorithms for recovering a 50-sparse signal $x \in I\!R^{512}$ from 120 measurements

Following the definition of the subdifferential set $\partial c(x_k)$ (2.3), we choose subgradient $\xi_k \in \partial c(x_k)$ as

$$(\xi_k)_i = \begin{cases} 1, & (x_k)_i > 0, \\ 0, & (x_k)_i \neq 0, \\ -1, & (x_k)_i < 0. \end{cases} \tag{3.9}$$

This algorithm, Algorithm 3, is denoted in our plots (Figs. 1 and 2) as Mod CQ (l1-con.).

Our schemes, DC (difference of convex) algorithm (DCA)-iterative step (2.11), the forward-backward (FB) algorithm-iterative step (2.45), and the Mine and Fukushima algorithm-Algorithm 2.3 are denoted in our plots (Figs. 1 and 2) as DCA (l_1-l_2), FB and Mine and Fukushima, respectively. The stopping criterion for all schemes is either 1000 iterations or until $\|x_{k+1} - x_k\| < 10^{-5}$ is reached. In the experiments, we choose arbitrary the regularization parameter γ to be 0.6. We noticed that this choice produces good results, and this also affects the sensitivity of the solution. All computations were performed using MATLAB R2015a on an Intel Core i5-4200U 2.3GHz running 64-bit Windows.

Next, the two numerical illustrations are presented. In Figs. 1 and 2, we present the performances of our schemes as well as the CQ and the modified CQ algorithms for random data and sparse signal recovering, respectively. As explained above, the algorithms are designed to solve $Ax = b$ with and without different types of regularizations. In Fig. 1, we present the results for the 50 random generated problems; in each plot, the different colors represent the quintiles with respect to each of the 50 problems and the red graph is the experiment median. It can be seen that most methods differ in their "warmup" stages, that is, in the first number of iterations, and all converge "quite fast," just within a few iterations. We see that the "warmup" stage in the DCA is the most significant and visible; we suspect that this is probably due to the need to solve subproblems during each iteration. This would probably play an essential role as a computational aspect for large-scale problems. In Fig. 2, we test the five scheme performances for recovering a 50-sparse signal $x \in \mathbb{R}^{512}$ from 120 measurements. Here, when only the resulting recovered signal is presented, it can be seen that the DCA, FB, and Mine and Fukushima algorithms recover the exact signal while both the CQ and the modified CQ algorithms contain errors, and as expected, the modified CQ algorithm generates a slightly better signal, probably due to the l1 -regularization. We would like to emphasize that the main goal of this work is to introduce and survey some approaches for solving $Ax = b$ with different variants of regularizations; we dont wish to further investigate and analyze the computational performances of the proposed schemes and hence wish to leave our above explanations as compact as possible. An interesting direction for future study is indeed a computational comparison between different types of regularizations. We believe that deep insights in this case can be derived, only when large problems are considered; since then, the subproblems solved per each iteration in the related algorithms might play an essential role with respect to the computational efforts and convergence rate, and moreover, this could emphasize and suggest the applicability and advantages of the different methods and in particular the usage of one regularization over another.

4 Concluding remarks

In this paper, we investigate split feasibility problems under a nonconvex Lipschitz continuous metric instead of conventional methods such as l_1 or $l_1 - l_2$ minimization, for example in [1]. We present and analyze the convergence to a stationary point of an iterative minimization method based on DCA (difference of convex algorithm); see for example [19]. Furthermore, relying on a proximal operator for $l_1 - l_2$ as well as on an algorithm proposed by Mine and Fukushima for minimizing the sum of a convex function and a differentiable one, two additional algorithms are presented and their convergence properties are discussed.

Since each iteration of the DCA requires to solve an inner l_1 -regularized split feasibility subproblem, we present some algorithms designed for that purpose in the Appendix. Observe that the DCA presented here can be extended to split fixed-point problems governed by firmly quasi-nonexpansive mappings. We would also like to emphasize that much attention has been paid not only to the sparsity of solutions but also to the structure of this sparsity, which may be relevant in some problems and which provides another avenue for inserting prior knowledge into the problem.

We would like to mention that an interesting regularizer is the OSCAR one which has the following form:

$$r_{OSCAR}(x) = \gamma_1 \|x\|_1 + \gamma_2 \sum_{i<j} \max\{|x_i|, |x_j|\}. \tag{4.1}$$

Due to l_1 term and the pairwise l_∞ penalty, the components are encouraged to standard spare and pairwise similar magnitude, have been extensively applied in various feature grouping tasks, and outperform other models. We refer to the interesting paper [25] where the OSCAR regularizer is used via its proximity mapping, a work that deserves to be more developed.

Acknowledgments We wish to thank the anonymous referee for the thorough analysis and review; all the comments and suggestions helped tremendously in improving the quality of this paper and made it suitable for publication. The first author wish to thank his team "Image & Modèle" and the Computer Science and System Laboratory (L.I.S) of Aix-Marseille University. The second author work is supported by the EU FP7 IRSES program STREVCOMS, grant no. PIRSES-GA-2013-612669.

Appendix A

Each DCA iteration requires solving a l_1 -regularized split feasibility subproblem of the form

$$\min_{x \in C} \frac{1}{2} \|(I - P_Q)Ax\|_2^2 + \langle x, v\rangle + \gamma \|x\|_1, \tag{A.1}$$

where $v \in I\!R^n$ is a constant vector. This problem can be done, for example, by the two split proximal algorithms (coupling the forward-backward and Douglas-Rachford algorithms).

A.1 Insertion of a forward-backward step in the Douglas-Rachford algorithm

To apply the Douglas-Rachford algorithm when $g_1 = \gamma\| \cdot \|_1$ and $g_2 = \frac{1}{2}\|(I - P_Q)A(\cdot)\|_2^2 + \langle \cdot, v\rangle + i_C$, we need to determine their proximal mappings. The main difficulty lies in the computation of the second one, namely $prox_{\kappa \frac{1}{2}\|(I-P_Q)A(\cdot)\|_2^2 + \langle \cdot, v\rangle + i_C}$. As in [9], we can use a forward-backward algorithm to achieve this goal.

The resulting algorithm is as follows:

Algorithm:

Step 1. Set $\underline{\gamma} \in]0, 2\kappa^{-1}\|A\|^{-1}]$, $\underline{\lambda} \in]0, 1]$ and $\kappa \in]0, +\infty[$.
Choose $(\tau_k)_{k\in I\!N}$ satisfying $\forall k \in I\!N,\ \tau_k \in]0, 2[$ and $\sum_{k=0}^{\infty} \tau_k(2 - \tau_k) = +\infty$

Step 2. Set $k = 0$, $y_0 = y_{-1/2} \in C$

Step 3. Set $x_{k,0} = y_{k-1/2}$

Step 4. For $i = 0, \ldots, N_k - 1$

a) Choose $\gamma_{k,n} \in [\underline{\gamma}, 2\kappa^{-1}\|A\|^{-1}[$ and $\lambda_{k,i} \in [\underline{\lambda}, 1]$.;
b) Compute

$$x_{k,i+1} = x_{k,i} + \lambda_{k,i}\left(P_C(\frac{x_{k,i} - \gamma_{k,i}(\kappa(A^t(I - P_Q)Ax_{k,i} + v_i) - y_k)}{1 + \gamma_{k,i}}) - x_{k,i}\right). \tag{A.2}$$

Step 5. Set $y_{k+1/2} = x_{k,N_k}$
Step 6. Set $y_{k+1} = y_k + \tau_k(prox_{\kappa\|\cdot\|_1}(2y_{k+1/2} - y_k) - y_{k+1/2})$.
Step 7. Increment $k \leftarrow k+1$ and go to step 3.

By a judicious choose of N_k, the convergence of the sequence (y_k) to y such that

$$prox_{\kappa(\frac{1}{2}\|(I-P_Q)A(\cdot)\|_2^2+\langle\cdot,v\rangle)+i_C}(y) \tag{A.3}$$

solves problem (A.1) follows directly by applying [9, Proposition 4.1].

A.2 Insertion of a Douglas-Rachford step in the forward-backward algorithm

We consider $f_1 = \kappa\|\cdot\|_1 + i_C$ et $f_2 = \frac{1}{2}\|(I-P_Q)A(\cdot)\|_2^2 + \langle\cdot, v\rangle$. Since f_2 has a $\|A\|^2$-Lipschitz gradient, we can apply the forward-backward algorithm. This requires however to compute $prox_{i_C+\gamma_k\|\cdot\|}$ which can be performed with Douglas-Rachford iterations. The resulting algorithm is

Algorithm:

Step 1. Choose γ_k and λ_k satisfying assumptions $0 < \inf_k \gamma_k \le \sup_k \gamma_k < 2/\|A\|^2$, $0 < \underline{\lambda} \le \lambda_k \le 1$.
Set $\underline{\tau} \in]0, 2]$.
Step 2. Set $k = 0$, $x_0 \in C$
Step 3. Set $x'_k = x_k - \gamma_k(A^t(I - P_Q)Ax_n + v)$.
Step 4. Set $y_{k,0} = 2prox_{\gamma_k\|\cdot\|_1}x'_k - x'_k$.
Step 5. For $i = 0, \ldots, M_k - 1$.

a) Compute

$$y_{k,i+1/2} = P_C\left(\frac{y_{k,i} + x'_k}{2}\right) \tag{A.4}$$

b) Choose $\tau_{k,i} \in [\underline{\tau}, 2]$.
c) Compute $y_{k,i+1} = y_{k,i} + \tau_{k,i}(prox_{\gamma_k\|\cdot\|_1}(2y_{n,i+1/2} - y_{k,i}) - y_{n,i+1/2})$.
d) If $y_{k,i+1} = y_{k,i}$, then goto step 6.

Step 6. Set $x_{k+1} = x_k + \lambda_k(y_{k,i+1/2} - x_k)$.
Step 7. Increment $k \leftarrow k+1$ and go to step 3.

A direct application of [9, Proposition 4.2] ensures the existence of positive integers $(\overline{M}_k)$ such that if $\forall k \ge 0$ $M_k \ge \overline{M}_k$, then the sequence (x_k) weakly convergences to a solution of problem (A.1).

Remark A.1 Other split proximal algorithms may be designed by combining the fixed-point idea to compute the composite of a convex function with a linear operator introduced in [16] and the analysis developed for computing the proximal mapping of the sum of two convex functions developed in [9] and [18]. Primal-dual algorithms considered in [12] can also be used. Note that there are often several ways to assign the functions of (A.1) to the terms used in the generic problem.

References

1. Alghamdi, M.A., Ali Alghamdi, M., Shahzad, N., Naseer, H.-K.X.: Properties and Iterative Methods for the Q-Lasso, Abstract and Applied Analysis. Article ID 250943, 8 pages (2013)
2. Byrne, C.: Iterative oblique projection onto convex sets and the split feasibility problem. Inverse Prob. **18**, 441–453 (2002)
3. Byrne, C.: A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Prob. **20**, 103Â–120 (2004)
4. Censor, Y., Bortfeld, T., Martin, B., Trofimov, A.: A unified approach for inversion problems in intensity-modulated radiation therapy. Phys. Med. Biol. **51**, 2353–2365 (2006)
5. Censor, Y., Elfving, T.: A multiprojection algorithm using Bregman projections in a product space. Numer. Algorithms **8**, 221–239 (1994)
6. Censor, Y., Gibali, A., Lenzen, F., Schnorr, Ch.: The implicit convex feasibility problem and its application to adaptive image denoising. J. Comput. Math. **34**, 610–625 (2016)
7. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process Lett. **14**, 707–710 (2007)
8. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. SIAM J. Comput. **20**, 33–61 (1998)
9. Chaux, C., Pesquet, J.-C., Pustelnik, N.: Nested iterative algorithms for convex constrained image recovery problems. SIAM J. Imag. Sci. **2**, 730–762 (2009)
10. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20**, 33–61 (1998)
11. Combettes, P.L., Pesquet, J.-C.: A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE J. Sel. Top. Sign. Proces. **1**, 564–574 (2007)
12. Condat, L.: A generic proximal algorithm for convex optimization: application to total variation minimization. IEEE Signal Process Lett. **21**, 985–989 (2014)
13. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory **52**, 1289–1306 (2006)
14. Esser, E., Lou, Y., Xin, J.: A method for finding structured sparse solutions to non-negative least squares problems with applications. SIAM J. Imag. Sci. **6**, 2010–2046 (2013)
15. Lou, Y., Yan, M.: Fast $l_1 - l_2$ Minimization via a proximal operator. arXiv:1609.09530 (2017)
16. Micchelli, Ch.A., Shen, L., Xu, Y., Zeng, X.: Proximity algorithms for the L_1/TV image denoising model. Adv. Comput. Math. **38**, 401–426 (2013)
17. Mine, H., Fukushima, M.: A minimization method for the sum of a convex function and a continuously differentiable function. J. Optim. Theory Appl. **33**, 9–23 (1981)
18. Moudafi, A.: About proximal algorithms for Q-Lasso, Thai Mathematical Journal (2016)
19. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**, 227–234 (1995)
20. Qu, B., Xiu, N.: A note on the CQ algorithm for the split feasibility problem. Inverse Prob. **21**, 1655–1665 (2005)
21. Tang, Y.-C., Liu, L.-W., Gibali, A.: Note on the modified relaxation CQ algorithm for the split feasibility problem. Optim. Lett., 1–14. https://doi.org/10.1007/s11590-017-1148-3 (2017)
22. Tibshirani, R.: Regression Shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B **58**, 267–288 (1996)
23. Yin, P., Lou, Y., He, Q., Xin, J.: Minimization of l_{1-2} for compressed sensing. SIAM J. Sci. Comput. **37**, 536–563 (2015)
24. Xu, Z., Chang, X., Xu, F., Zhang, H.: L_{1-2} regularization: a thresholding representation theory and a fast solver. IEEE Trans. Neural Netw. Learn. Syst. **23**, 1013–1027 (2012)
25. Zeng, X., Figueiredo, M.A.-T.: Solving OSCAR regularization problems by fast approximate proximal splittings algorithms. Digital Signal Process. **31**, 124–135 (2014)