

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

A nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

A nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging / Lazzaro, D; Piccolomini, E Loli; Zama, F. - In: INVERSE PROBLEMS. - ISSN 0266-5611. - ELETTRONICO. - 35:8(2019), pp. 084002.1-084002.17. [10.1088/1361-6420/ab1c6b]

*Availability:*

This version is available at: <https://hdl.handle.net/11585/698194> since: 2019-09-11

*Published:*

DOI: <http://doi.org/10.1088/1361-6420/ab1c6b>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

**Lazzaro, D., E. L. Piccolomini, and F. Zama. "A Nonconvex Penalization Algorithm with Automatic Choice of the Regularization Parameter in Sparse Imaging." *Inverse Problems*, vol. 35, no. 8, 2019.**

The final published version is available online at : <http://dx.doi.org/10.1088/1361-6420/ab1c6b>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# An nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging

D. Lazzaro <sup>1</sup>, E. Loli Piccolomini <sup>2</sup>, F. Zama <sup>1</sup>

<sup>1</sup> Department of Mathematics, University of Bologna, Piazza di Porta San Donato, 5, 40126, Bologna (ITALY)

<sup>2</sup> Computer Science and Engineering Department, University of Bologna, Mura Anteo Zamboni, 7, 40126, Bologna (ITALY)

**Abstract.** This paper introduces an efficient method for solving nonconvex penalized minimization problems. The topic is relevant in many imaging problems characterized by sparse data. The proposed method originates from the Iterative Reweighting  $l_1$  scheme, modified by the automatic update of the regularization parameter on the basis of the behavior of the objective function. Besides proving the convergence of the method, a modified algorithm is obtained and the performance is tested on two different sparse imaging problems. The proposed method can be viewed as a general framework which can be adapted to different one-parameter non convex penalty functions and applied to problems characterized by sparse data.

## 1. Introduction

Nonconvex penalized minimization occurs in a wide variety of imaging and computer vision applications. The variational formulation of such problems involves a nonconvex objective function which can be expressed as follows:

$$\min_U J(U) + \lambda F(U) \tag{1}$$

where  $U$  represents the unknown image,  $J(U)$  is a, usually convex, fit-to-data function,  $F(U)$  is the so called penalization or regularization function and  $\lambda > 0$  is the unknown regularization parameter.

The choice of the regularization parameter is always critical and it usually involves heuristic procedures where the value is determined after many trials. Moreover, since a lower contribution of the regularization term is required in the iterative methods when approaching the solution, a strategy decreasing  $\lambda$  during the iterations is preferred.

Nonconvex penalties received increasing attention for their good properties in treating sparsity in the data. The excellent results obtained in the first works [1–3] encouraged the use of nonconvex penalties in a great number of imaging applications such as denoising, deblurring, super-resolution or sparse Compressive Sensing (CS)

reconstructions in medical MRI [4, 5] or X-rays Computed Tomography (CT)[6]. Frequently in imaging  $F$  depends on the image gradient  $DU$  that well captures the contrast changes in the image.

It is well known that the  $\ell_0$  quasi-norm, measuring the cardinality of its argument, is the best possible sparsifying function, but its minimization is computationally very expensive. A remarkable result of the CS theory [7, 8] states that the signal recovery is still possible if one substitutes the  $\ell_1$  norm to the  $\ell_0$  quasi-norm under suitable hypotheses; in order to better approximate the  $\ell_0$  quasi-norm, the  $\ell_p$ ,  $0 < p < 1$ , non-convex sparsifying norm can be used [4, 9]. Even if the global minimum of (1) cannot be guaranteed in this case, the authors present a proof of asymptotic convergence of  $\ell_p$  towards  $\ell_0$  by suitably setting the restricted isometry constants. In [5, 10] different non-convex families of sparsifying functions depending on a positive parameter and homotopic with the  $\ell_0$  seminorm are considered with very promising results.

The drawback of nonconvex minimization is the difficulty in numerically solving the optimization problem. Lately, many efforts have been done in proposing new efficient strategies or in adapting to the solution of problem (1) when  $F$  is non-convex the algorithms considered for convex optimization. The authors in [11] present an overview of the progress in non-convex optimization by considering algorithms belonging to the proximal descent and ADMM classes. Another approach, widely applied in imaging applications, for the solution of (1) is Iterative Reweighting, where the nonconvex penalty function is iteratively approximated by its convex linearization (see, for example, [4, 12, 13]).

**Paper contribution.** Aim of this paper is to propose a fast method with automatic choice of the regularization parameter  $\lambda$  for the solution of the penalized nonconvex problem (1).

The proposed method originates from the Iterative Reweighting  $l_1$  (IRl1) scheme [14] modified by a decreasing update of the regularization parameter  $\lambda$  at each iteration. We prove the convergence of this variational method named Variable Parameter Iterative Reweighting (VPIR). In particular, we use an adaptive rule, which decreases, at each iteration of the algorithm, the value of  $\lambda$  on the basis of the behavior of the objective function, as suggested in [15]. The inner convex problem obtained at each IRl1 iteration is convex and non differentiable. Several methods can be used for its solution such as Chambolle Pock [16], Split-Bregman [17], Alternating Minimization [18]. In this paper we use the Forward-Backward algorithm [19] since it requires the tuning of very few parameters, that is a great advantage in real applications. In particular in [20] we defined an Accelerated Forward Backward algorithm (AFB) and proved its efficiency in the solution of problems like (1).

Finally, we propose MVPIR, a Modified version of VPIR, which exhibits better computational efficiency.

The proposed methods, here applied with a particular choice of the non-convex regularization function  $F$ , can be viewed as general framework which can be adapted to different non-convex regularization functions in problems characterized by sparse data.

We report the results obtained in several numerical tests on two important imaging problems characterized by sparse data: Magnetic Resonance (MR) Image reconstruction from low-sampled data and image deblurring of QR-code images. The reconstruction of MR images from a reduced set of acquisitions is crucial to speed up the measurements in order to decrease the costs and improve the patient care [4, 5, 21, 22]. On the other hand, image deblur is a task required by many imaging applications and widely studied in the literature. The tests show a very good performance of VPIR method, improved by MVPIR for what concerns the computational coat.

The paper is organized as follows. In section 2 we present the nonconvex variational problem, in section 3 we analyze the VPIR and MVPIR methods and prove the convergence. Finally in section 4 we show the results obtained in the numerical experiments and at last in section 5 we report some conclusions.

## 2. The nonconvex regularization problem

In this section we discuss the properties of the nonconvex optimization problem (1) for the two imaging applications used to test our methods. The fit-to-data function  $J$  is the likelihood between the observed and the predicted data and depends on the probability distribution of the noise. Under the hypothesis that the noise has Gaussian distribution we model  $J$  by the least squares function:

$$J(u) = \frac{1}{2} \|\Phi u - z\|_2^2 \quad (2)$$

where  $u = \text{vec}(U)$  is the vector obtained by rearranging the elements of the image  $U$ ,  $z$  is the vector of the acquired data and  $\Phi$  is the matrix representing the imaging system.

In the sparse MRI application,  $z$  is a vector in the Fourier space (K-space) and  $\Phi$  is the under-sampling Fourier matrix, obtained by the Hadamard product between the full resolution Fourier matrix  $\mathbf{F}$  and the mask  $\mathcal{M}$ , i.e.

$$\Phi = \mathcal{M} \circ \mathbf{F}. \quad (3)$$

When the MR K-space is only partially measured the inversion problem has infinite possible solutions. Following the CS theory, a way to choose one of these infinite solutions is to impose a prior in a sparsity domain.

In image deblur,  $\Phi$  is the discretization of the Point Spread function (PSF) describing the blur produced by the image acquisition instrument, such as lens, microscope, telescope or camera, while  $z$  is the blurred image.

Concerning the choice of the regularization or sparsifying function  $F(DU)$ , we define a class of non-convex functions  $F_\mu(DU)$ , depending on the parameter  $\mu > 0$ , satisfying the property:

$$\lim_{\mu \rightarrow 0} \sum_{\Omega} F_\mu(DU) = \|DU\|_0. \quad (4)$$

where  $\Omega$  is the image domain,  $DU = (U_x, U_y)$  is the discrete gradient of the image  $U$  and

$$F_\mu(DU) = \sum_{i,j} (\psi_\mu(|U_x|_{i,j}) + \psi_\mu(|U_y|_{i,j})). \quad (5)$$

In our experiments we choose the function  $\psi_\mu(t) : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$  having the following expression:

$$\psi_\mu(t) = \frac{1}{\log(2)} \log \left( \frac{2}{1 + e^{-\frac{|t|}{\mu}}} \right), \quad \mu > 0 \quad (6)$$

that has been shown in [10] to be very well performing in a CS framework. In the same paper the authors show that  $\psi_\mu(t)$  has the property

$$\lim_{\mu \rightarrow 0} \psi_\mu(t) = \|t\|_0$$

and that it is an increasing, nonconvex, symmetric, twice differentiable function.

### 3. The method

In this section, we consider the non convex minimization problem:

$$\min_u \left\{ \lambda F_\mu(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\}, \quad \mu > 0. \quad (7)$$

and we propose the VPIR method based on the IRl1 scheme [14] with a variable regularization parameter update for each iteration. We report the steps of the algorithm VPIR) in Table 1.

Starting from an initial guess  $\bar{u}^{(0)}$ , the IRl1 method computes a sequence of approximate solutions  $\{\bar{u}^{(h)}\}$ ,  $h = 1, 2, \dots$ , by substituting convex approximations to the original objective function in (1).

In particular, at each iteration  $h$  of the IRL1 algorithm, the nonconvex function  $F_\mu(DU)$  is locally approximated by its linear convex majorizer  $\mathcal{F}_{h,\mu}(DU)$ :

$$\mathcal{F}_{h,\mu}(DU) = F_\mu(D\bar{U}^{(h)}) + f_{h,\mu}(DU) - f_{h,\mu}(D\bar{U}^{(h)}) \quad (9)$$

where

$$f_{h,\mu}(DU) = \sum_{i,j=1}^N (\psi'_\mu(|\bar{U}_x^{(h)}|_{i,j})|U_x|_{i,j} + \psi'_\mu(|\bar{U}_y^{(h)}|_{i,j})|U_y|_{i,j}) \quad (10)$$

with  $\psi(t)$  defined by (6). Hence, at each iteration  $h$  we solve the convex minimization problem:

$$\bar{u}^{(h)} = \arg \min_u \mathcal{P}(\lambda, \mathcal{F}_{h-1,\mu}(DU), u) = \arg \min_u \left\{ \lambda \mathcal{F}_{h,\mu}(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\} \quad (11)$$

Since  $f_{h,\mu}(DU)$  is the only term of (9) depending on  $U$ , the minimization problem (11) reduces to the following convex minimization problem:

$$\bar{u}^{(h)} = \arg \min_u \left\{ \lambda f_{h-1,\mu}(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\}. \quad (12)$$

---

VPIR algorithm
(Input: $z, r_0, \eta, \Phi$ ) $\bar{u}^{(0)} = \phi^T z, \lambda_0 = r_0 \ u^{(0)}\ _1, \mu = \ \Delta u^{(0)}\ _1$ $w_{i,j}^x = 1, w_{i,j}^y = 1, \forall i, j = 1, \dots, N$ $k = 0;$ <b>repeat</b> $h = 0$ <b>repeat</b> (IRl1 algorithm) I.1 Compute $\bar{u}^{(h+1)}$ in (8) starting from $\bar{u}^{(0)}$ $\bar{u}^{(h+1)} = \arg \min_u \left\{ \lambda_h f_\mu(Du) + \frac{1}{2} \ \Phi u - z\ _2^2 \right\} \quad (8)$ where : $f_\mu(Du) = \sum_{i,j} (w_{i,j}^x  \bar{u}_x _{i,j} + w_{i,j}^y  \bar{u}_y _{i,j})$ I.2 Compute the weights: $w_{i,j}^x = \psi'_\mu( \bar{u}_x^{(h+1)} _{i,j}), w_{i,j}^y = \psi'_\mu( \bar{u}_y^{(h+1)} _{i,j})$ I.3 Update $\lambda_{h+1}$ <b>if</b> $h=0$ <b>then</b> $\lambda_1 = \lambda_0/2$ <b>else</b> update $\lambda_{h+1}$ as in (19) I.4 $h = h + 1$ <b>until</b> convergence $k = k + 1,$ $\mu = \mu \cdot \eta$ Continuation strategy $u^{(k)} = \bar{u}^{(h)}$ $\bar{u}^{(0)} = u^{(k)}$ Warm starting <b>until</b> stopping rule

---

**Table 1.** VPIR Algorithm

We propose here to modify the value of the regularization parameter  $\lambda$  at each iteration by defining a decreasing sequence  $\{\lambda_h\}$ .

Hence the VPIR algorithm computes a sequence  $\{\bar{u}^{(h)}\}$ ,  $h = 1, 2, \dots$  where:

$$\bar{u}^{(h)} = \arg \min_u \left\{ \lambda_{h-1} f_{h-1, \mu}(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\}. \quad (13)$$

We prove that the iterates  $\bar{u}^{(h)}$ , computed by (13), converge to a local minimum of problem (7).

To this purpose we define:

$$\mathcal{P}(\lambda, F_\mu(DU), u) = \left\{ \lambda F_\mu(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\}. \quad (14)$$

By using (9) we have:

$$\mathcal{F}_{h-1,\mu}(D\bar{U}^{(h-1)}) = F_\mu(D\bar{U}^{(h-1)}) \quad \text{and} \quad \mathcal{F}_{h-1,\mu}(DU) \geq F_\mu(DU) \quad u \neq \bar{u}^{(h-1)}. \quad (15)$$

The following result proves the descent property of the objective function at the solution computed by (13).

**Proposition 3.1.** *Let  $\bar{\mu}$  be a fixed value of  $\mu$  and  $\lambda_{h-1}$ ,  $\lambda_h$  be two values of the penalization parameter such that  $\lambda_{h-1} > \lambda_h$ . Let  $\bar{u}^{(h)}$  and  $\bar{u}^{(h+1)}$  be the corresponding minimizers given by (11) with  $\lambda = \lambda_{h-1}$  and  $\lambda = \lambda_h$ , respectively. Then the nonconvex functional  $\mathcal{P}(\lambda, F_{\bar{\mu}}(DU), u)$  satisfies the following inequality:*

$$\mathcal{P}(\lambda_{h-1}, F_{\bar{\mu}}(D\bar{U}^{(h)}), \bar{u}^{(h)}) > \mathcal{P}(\lambda_h, F_{\bar{\mu}}(D\bar{U}^{(h+1)}), \bar{u}^{(h+1)})$$

**Proof:** Using relation (3) and rewriting the first part of property (15) for  $h$ :

$$\mathcal{F}_{h,\mu}(D\bar{U}^{(h)}) = F_\mu(D\bar{U}^{(h)}),$$

we have:

$$\begin{aligned} \mathcal{P}(\lambda_{h-1}, F_{\bar{\mu}}(D\bar{U}^{(h)}), \bar{u}^{(h)}) &= \frac{1}{2} \|\Phi \bar{u}^{(h)} - z\|_2^2 + \lambda_{h-1} F_{\bar{\mu}}(D\bar{U}^{(h)}) \\ &= \frac{1}{2} \|\Phi \bar{u}^{(h)} - z\|_2^2 + \lambda_{h-1} \mathcal{F}_{h,\bar{\mu}}(D\bar{U}^{(h)}) \end{aligned}$$

From the assumption  $\lambda_{h-1} > \lambda_h$  it follows that:

$$\frac{1}{2} \|\Phi \bar{u}^{(h)} - z\|_2^2 + \lambda_{h-1} \mathcal{F}_{h,\bar{\mu}}(D\bar{U}^{(h)}) > \frac{1}{2} \|\Phi \bar{u}^{(h)} - z\|_2^2 + \lambda_h \mathcal{F}_{h,\bar{\mu}}(D\bar{U}^{(h)})$$

Using (11):

$$\frac{1}{2} \|\Phi \bar{u}^{(h)} - z\|_2^2 + \lambda_h \mathcal{F}_{h,\bar{\mu}}(D\bar{U}^{(h)}) \geq \frac{1}{2} \|\Phi \bar{u}^{(h+1)} - z\|_2^2 + \lambda_h \mathcal{F}_{h,\bar{\mu}}(D\bar{U}^{(h+1)})$$

Applying the second part of property (15) to the case  $DU \equiv D\bar{U}^{(h+1)}$ :

$$\frac{1}{2} \|\Phi \bar{u}^{(h+1)} - z\|_2^2 + \lambda_h \mathcal{F}_{h,\bar{\mu}}(D\bar{U}^{(h+1)}) \geq \frac{1}{2} \|\Phi \bar{u}^{(h+1)} - z\|_2^2 + \lambda_h F_{\bar{\mu}}(D\bar{U}^{(h+1)})$$

From definition (3):

$$\frac{1}{2} \|\Phi \bar{u}^{(h+1)} - z\|_2^2 + \lambda_h F_{\bar{\mu}}(D\bar{U}^{(h+1)}) = \mathcal{P}(\lambda_h, F_{\bar{\mu}}(D\bar{U}^{(h+1)}), \bar{u}^{(h+1)})$$

hence:

$$\mathcal{P}(\lambda_{h-1}, F_{\bar{\mu}}(D\bar{U}^{(h)}), \bar{u}^{(h)}) > \mathcal{P}(\lambda_h, F_{\bar{\mu}}(D\bar{U}^{(h+1)}), \bar{u}^{(h+1)})$$

This proves the result. □

In order to prove the convergence we need the following results.



**Definition 3.1.** A convex (not necessarily differentiable) function  $f(u)$  is said  $\nu$ -strongly convex if and only if there exists a constant  $\nu > 0$ , called the modulus of strong convexity of  $f(u)$ , such that the function  $f(u) - \frac{\nu}{2}\|u\|_2^2$  is convex.

**Proposition 3.2.** The functional  $\mathcal{P}(\lambda_h, \mathcal{F}_{h,\bar{\mu}}(DU), u)$  is  $\nu$ -strongly convex with  $\sigma = \frac{1}{\lambda_h}$ .

The proof of Proposition 3.2 follows straightforward by observing that  $\mathcal{P}(\lambda_h, \mathcal{F}_{h,\bar{\mu}}(DU), u)$  is the sum of a  $\nu$ -strongly convex term (the  $\ell_2$ -norm is  $\nu$ -strongly convex with  $\nu = 1$ ) and of a convex term, due to the  $\ell_1$ -norm. It follows that is  $\nu$ -strongly convex, see [23].

**Proposition 3.3.** Let's define the bounding set  $B$  s.t.  $\bar{u}^{(h)} \in B$ ,  $\forall h \geq 0$ , let's assume that  $F_{\bar{\mu}}$  and  $\mathcal{F}_{h,\bar{\mu}}$  have a locally Lipschitz continuous gradient on  $B$  with a common Lipschitz constant  $\tilde{L} \geq 0$ . Then the following properties hold:

(i) Descent property:

$$\mathcal{P}(\lambda_{h-1}, F_{\bar{\mu}}(D\bar{U}^{(h)}), \bar{u}^{(h)}) > \mathcal{P}(\lambda_h, F_{\bar{\mu}}D\bar{U}^{(h+1)}), \bar{u}^{(h+1)}) \quad (16)$$

(ii) There exists  $C > 0$  such that for all  $h \in N$  there exists

$$\xi_{h+1} \in \partial \mathcal{P}(\lambda_h, F_{\bar{\mu}}(D\bar{U}^{(h+1)}), \bar{u}^{(h+1)})$$

fulfilling

$$\|\xi_{h+1}\|_2 \leq C \|\bar{u}^{(h+1)} - \bar{u}^{(h)}\|_2 \quad (17)$$

(iii) For any converging subsequence  $\{\bar{u}^{(h_j)}\}_{j \in N}$  with

$$\bar{\bar{u}} \equiv \lim_{j \rightarrow \infty} \bar{u}^{(h_j)},$$

and  $\{\lambda_{h_j}\}$  with  $\bar{\bar{\lambda}} = \lim_{j \rightarrow \infty} \lambda_{h_j}$ , we have

$$\mathcal{P}(\lambda_{h_j}, F_{\bar{\mu}}(D\bar{U}^{(h_j)}), \bar{u}^{(h_j)}) \rightarrow \mathcal{P}(\bar{\bar{\lambda}}, F_{\bar{\mu}}(D\bar{\bar{u}}), \bar{\bar{u}}) \quad j \rightarrow \infty. \quad (18)$$

**Proof:** Using Proposition 3.1 we can prove the descent property (16). Relations (17) and (18) can be easily proved as in [13], proposition 5.  $\square$

**Proposition 3.4.** Let  $\bar{\mu} > 0$  be a given value then the sequence  $\bar{u}^{(h)}$  defined in (13) converges to a local minimum of problem (1).

**Proof:** We observe that the objective function

$$\left\{ \lambda_h F_{\bar{\mu}}(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\}, \quad \lambda_h > 0.$$

satisfies the Kurdika-Lojasiewicz property [24]. In fact it is an analytic function, because the function  $\psi$  in (10) is evaluated in nonnegative arguments, therefore we can restrict the domain of  $\psi$  on  $R^+ \cup \{0\}$ .

Finally the convergence immediately follows from [13], where the authors prove that the

Iterative Reweighted Algorithm converges, provided that the relations (16), (17), (18) hold and the objective function satisfies the Kurdika-Lojasiewicz property.  $\square$

A possible updating rule for  $\lambda_h$  is given in [15], where, starting from a sufficiently large initial estimate  $\lambda_0$  and choosing  $\lambda_1 < \lambda_0$ , we compute:

$$\lambda_{h+1} = \lambda_h \frac{\mathcal{P}(\lambda_h, \mathcal{F}_{h,\mu}(D\bar{U}^{(h)}), \bar{u}^{(h)})}{\mathcal{P}(\lambda_{h-1}, \mathcal{F}_{h-1,\mu}(D\bar{U}^{(h-1)}), \bar{u}^{(h-1)})}, \quad h = 1, \dots \quad (19)$$

In [15] it is proven that the sequence  $\{\lambda^{(h)}\}$ ,  $h = 1, 2, \dots$  is strictly decreasing.

Following the idea proposed in [5] in order to enhance the sparsity in the gradient domain, we solve problem (13) for decreasing values of  $\mu$  and by using a continuation strategy for the initial iterate, as reported in Table 1.

Finally we propose a modified version of VPIR algorithm, obtained by inexactly solving the problem (13) where the  $h$  loop is stopped after only one iteration. We have heuristically chosen only one iteration on the basis of some numerical trials, but obviously the idea can be extended by stopping the IRl1 method after  $h^*$  iterations far before the convergence. We name this new algorithm as Modified VPIR (MVPIR). This modification can also be viewed as the elimination of the outer loop of VPIR and the computation of a sequence of approximate solutions by minimizing the following modified objective function:

$$\left\{ \lambda_h f_{\mu_h}(DU) + \frac{1}{2} \|\Phi u - z\|_2^2 \right\}$$

where the the penalty function  $f_\mu$  in (8) is substituted by  $f_{\mu_h}$  where  $\mu_h$  decreases at each IRl1 step. The steps of MVPIR are reported in Table 2.

MVPIR algorithm
(Input: $z, r_0, \eta, \phi$ )
$u^{(0)} = \phi^T z, \lambda_0 = r_0 \ u^{(0)}\ _1, \mu = \ \Delta u^{(0)}\ _1$
$w_{i,j}^x = 1, w_{i,j}^y = 1, \forall i, j = 1, \dots, N$
$k = 0$
<b>repeat</b>
Steps I.1- I.4 as in Algorithm VPIR (Table:1);
$\mu = \mu \cdot \eta$ <i>Continuation strategy</i>
$u^{(0)} = u^{(k)}$ <i>Warm starting</i>
<b>until stopping criterion</b>

**Table 2.** MVPIR Algorithm

Although the convergence of MVPIR algorithm is not proven, we can appreciate its efficiency on both noiseless and noisy data for different types of applications.

## 4. Numerical Results

The aim of this section is to analyze the characteristics and performance of VPIR and MVPIR. Two important imaging problems are considered: Magnetic Resonance Image Reconstruction from sparse acquisitions (paragraph 4.1), and Image Deblurring (paragraph 4.2).

In our tests we apply the algorithms to reconstruct images  $U$  from data possibly affected by noise. We evaluate the quality of the computed images  $U$  by means of the Signal to Noise Ratio (SNR) and/or the Peak Signal to Noise Ratio (PSNR) parameters:

$$SNR = 20 \log_{10} \frac{\sum_i \sum_j U_{i,j}^2}{\sum_i \sum_j (U_{i,j} - X_{i,j})^2}, \quad PSNR = 20 \log_{10} \frac{\max(X)}{\sqrt{\frac{\sum_i \sum_j (U_{i,j} - X_{i,j})^2}{N^2}}}.$$

where  $X$  represents the true reference image.

### 4.1. Sparse MRI experiments

In this paragraph we compare VPIR and MVPIR algorithms both on noiseless and noisy low-sampled MR data. Although the results well compare to the latest achievements, an accurate comparison with the recent literature (see, for example, [22, 25]) is outside the scope of this paper.

All the algorithms are implemented in Matlab R17b on a PC equipped with Xeon(R) processors 2.4 GHz and 256GB Ram.

The synthetic undersampled data  $z$  are obtained as  $z = \Phi u$  where  $u$  is the phantom image and  $\Phi$  is the undersampling Fourier matrix (3). In some experiments we test the algorithms on noisy data obtained by adding complex Gaussian white noise of levels  $\delta = 10^{-3}, 10^{-2}$  to the acquired sub-sampled data  $z$  as follows:

$$z^\delta = z + \delta \|z\| v, \quad \|v\| = 1 \quad (20)$$

where  $v$  is a complex unit norm random vector.

In the experiments presented in this section the mask  $\mathcal{M}$  in (3) is a radial subsampling mask with a variable percentage of nonzero elements (Figure 1. The test image is the famous **Shepp-Logan** phantom of size  $256 \times 256$  (figure 2(a)) which is widely used in algorithm testing. The high resolution images are obtained from the low-sampled data by Fourier inversion of the zero filled ( $zf$ ) data as well as by applying a gridding method before the Fourier inversion ( $gr$ ) using the fast Gaussian gridding **fNUFFT** [26], available in the Matlab file exchange.

We report in figure 2(b) the high resolution image  $u^{(0)} = \Phi^T z$  obtained, which is used as the starting point of both algorithms (mettere sia zero filling che gridding? con quale percentuale di dati?) In order to be consistent with the typical parameters used to represent the quality of MR Images, we also compute the Contrast to Noise Ratio (CNR) parameter by considering two Regions Of Interest (ROI) containing details (ROI

A and ROI B in Figure 2 (a)) and a ROI in the background (ROI C in Figure 2 (a)) and by using the mean and standard deviation values over the regions as:

$$CNR = \frac{\text{mean}(ROIA) - \text{mean}(ROIB)}{\text{std}(ROIC)}.$$

The stopping criteria used in both VIPR and MVPIR are based on the relative distance  $\tau$  between two successive iterations and on the decrease of the residual norm:

$$\|\bar{u}^{(h)} - \bar{u}^{(h-1)}\| \leq \tau \|\bar{u}^{(h)}\| \quad \text{OR} \quad \|\Phi \bar{u}^{(h)} - z\| > \|\Phi \bar{u}^{(h-1)} - z\| \quad (21)$$

The same criterion is used in VIPR to stop the outer iterations of the continuation method; in this case (21) is applied to  $u^{(k)}$ . In all the experiments the parameter  $\tau$  is equal to the machine precision. We observe that the condition on the residual acts on noisy data because in that case the residual norm loses its regular decreasing behavior. In case of noiseless data residual norm decreases and the small tolerance  $\tau$  allows us to reach very accurate solutions.

We report in Table 3 the PSNR, SNR and CNR parameters together with the computation time for both noiseless and the noisy data. We compare the VIPR and MVPIR results by using both the zero-filling (*zf*) *interpolation gridding* (*gr*) techniques for obtaining the high resolution image from the low resolution data in each step of the considered algorithms.

In case of noiseless data we observe that the two algorithms reach both very accurate results although MVPIR has slightly higher PSNR values. Comparing the computation times we have that MVPIR is about 60% faster. When noisy data are used, we observe from the tests reported in Table 3 that the two algorithms reach a comparable computation time while VIPR reaches slightly better results in terms of PSNR. The use of gridding does not improve the results in terms of image quality but increases the computational time, since the technique is applied several times (one each iteration of the methods).

Figure 2 compares some high reconstructions: in (b) the image simply obtained by zero-filling, in (c) the reconstruction performed by the NUFFT functions using gridding interpolation, in (d) and (e) the VIPR and MVPIR reconstructions (all the cases are from the 7.9% of data with noise  $\delta = 10^{-2}$ ). The images clearly show that the VIPR and MVPIR images have far better quality of the zero-filled and gridding interpolation reconstructions. Observing the error histogram reported in figure 5 we see that MVPIR has a slightly larger number of zeros and smaller values compared to VIPR.

The efficiency of the two algorithms can be appreciated in figures 3(a) and 3(b) where the PSNR versus time curves show a faster PSNR increase in MVPIR. The efficiency of the stopping rule and automatic computation of the regularization parameters  $\lambda$  is shown in the RMSE versus  $\lambda$  curves reported in figures 4. We observe that MVPIR stops at the minimum RMSE both in noisy and noiseless data. Concerning VIPR we observe a slight increase of the RMSE for noisy data which means that the value of the regularization parameter is slightly underestimated.

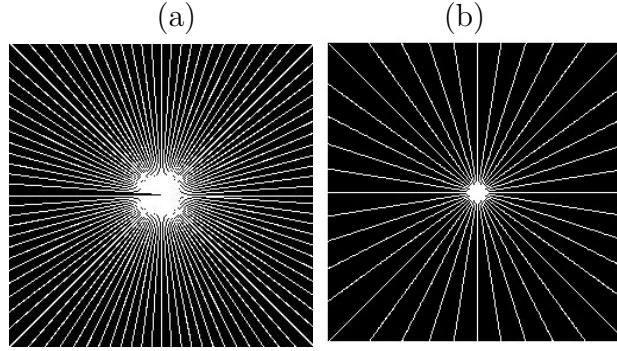
$Nz$	Method	$\delta$	PSNR		SNR		CNR		time(s)	
			$zf$	$gr$	$zf$	$gr$	$zf$	$gr$	$zf$	$gr$
20%	VPIR	0	315.26	321.46	309.78	310.4	$10^{13}$	$10^{13}$	11.8	160.25
		$10^{-3}$	64.30	66.54	52.13	54.37	59.63	82.55	2.77	7.61
		$10^{-2}$	48.19	48.38	34.25	35.03	8.60	10.24	2.3	6.84
	MVPIR	0	323.13	323.89	310.42	311.72	$10^{13}$	$10^{13}$	3.15	27.2
		$10^{-3}$	67.9	68.11	55.78	55.94	110.02	117.08	2.37	5.79
		$10^{-2}$	49.97	50.49	37.78	38.24	11.32	13.11	1.98	4.98
12%	VPIR	0	320.5	318.9	308.7	306.7	$10^{13}$	$10^{13}$	23.94	156.84
		$10^{-3}$	70.1	57.9	59.8	47.6	101.2	36.5	4.77	6.3
		$10^{-2}$	46.5	39.27	34.3	27.01	8.6	4.35	3.18	3.33
	MVPIR	0	323.3	324.1	311.2	312.0	$10^{13}$	$10^{13}$	4.84	41.78
		$10^{-3}$	66.9	65.2	54.7	53.0	70.6	59.7	2.10	6.84s
		$10^{-2}$	46.2	43.21	33.9	30.53	9.0	6.16	2.17	3.75
8%	VPIR	0	304.7	289.9	292.6	277.7	$10^{13}$	$10^{13}$	24.16	161.5
		$10^{-3}$	61.7	70.2	49.5	58.0	53.8	119.5	8.91	26.25
		$10^{-2}$	38.0	36.53	25.8	24.19	4.1	4.78	4.67	11
	MVPIR	0	310.1	315.2	298.0	302.9	$10^{13}$	$10^{13}$	7.03	65.34
		$10^{-3}$	59.4	57.2	47.2	45.3	49.9	35.12	2.35	8.29
		$10^{-2}$	28.4	28.51	15.9	15.96	1.8	1.83	1	3

**Table 3.** Sparse MRI experiments on Shepp Logan Phantom ( $256 \times 256$ ). Radial Undersampling Mask with non zero elements ( $Nz$ ) 20%, 12% and 8%.  $zf$  refers to Zero filling while  $gr$  refers to gridding.

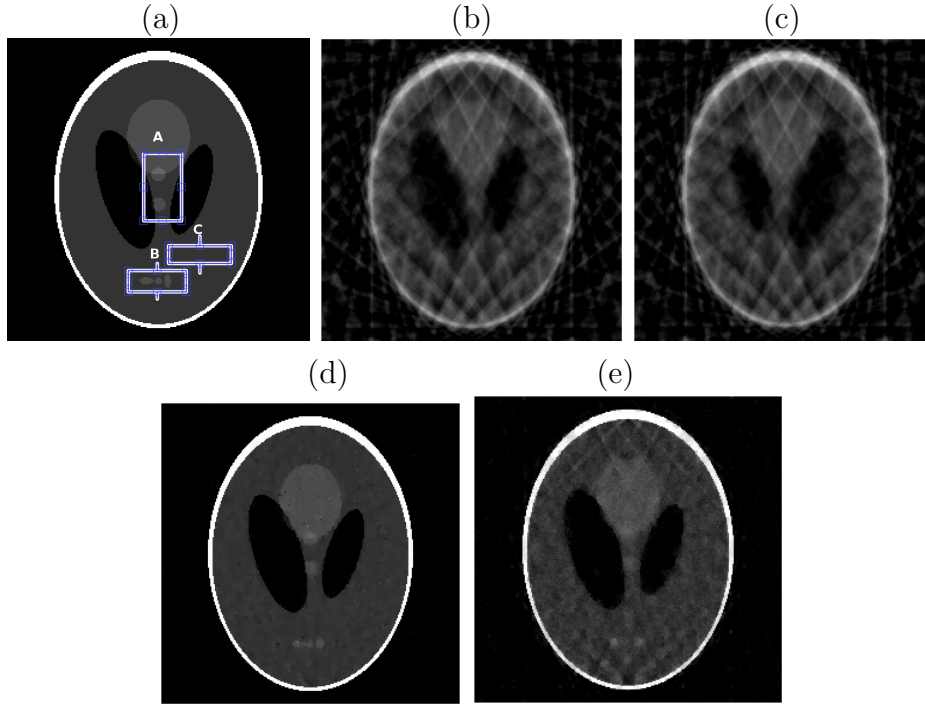
Finally a more accurate analysis can be made by running the algorithms on 100 different random noise realizations. The boxplots of the computed PSNR and computational time in seconds relative to these tests are reported in figures 6 and 7, respectively. We observe from Figure 6 that MVPIR has the best median values of PSNR especially for the highest noise ( $\delta = 10^{-2}$ ). Figure 7 confirms the better computational performance of MVPIR. Moreover the lower number of outliers (red crosses) and smaller extension of the boxes (blue rectangles) proves the best stability of MVPIR.

#### 4.2. Image deblur

In this paragraph we consider an image deblurring problem applied to a QR-code image, which is very sparse in the gradient domain. We consider here only the MVPIR implementation, since from the MRI experiments is evident that is it more advantageous than VPIR. Numerical results are presented to demonstrate the effectiveness of the proposed approach also compared with the L2-TV based image restoration where the Total Variation (TV) penalization function has been used instead of the nonconvex penalty in (1). In this case we solve problem (8) by using the TV function in place of the weighted TV deriving from the linearization of the nonconvex penalty.



**Figure 1.** Radial sampling Masks  $\mathcal{M}$ . (a) 20%, (b) 8% non zero elements.



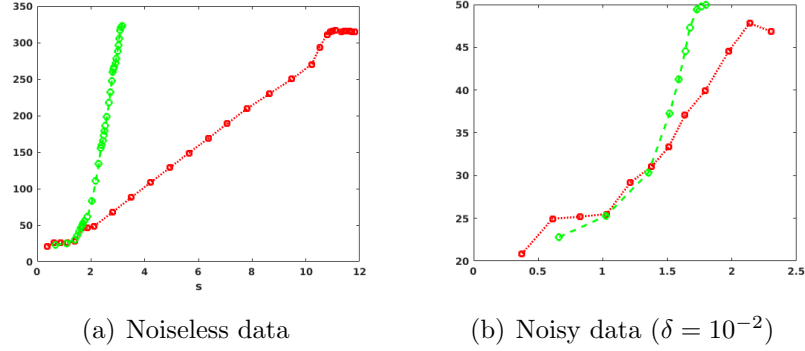
**Figure 2.** Shepp Logan Phantom ( $256 \times 256$ ). (a) Reference image. (b) zero-filled reconstruction (PSNR=17.6) (c) gridding interpolation reconstruction obtained with NUFFT functions (PSNR=17.2) (d) VPIR reconstruction (e) MVPIR reconstruction ( $Nz = 8\%$ ,  $\delta = 10^{-2}$ )..

The results presented in this paragraph are obtained by running a MATLAB implementation of MVPIR on a PC with Intel Core(TM) i7-3770 CPU (3.4 GHz) and 16 GB of RAM.

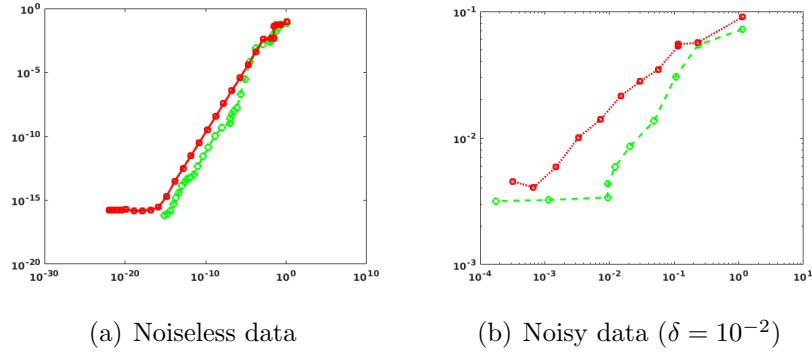
We measure the quality of the restored image by the PSNR value defined in (??) and the amount of blur and noise in the observed image with the BSNR value defined as follows:

$$BSNR = 20 \log_{10} \left( \frac{\|z - \bar{z}\|}{\sigma_n \sqrt{N}} \right)$$

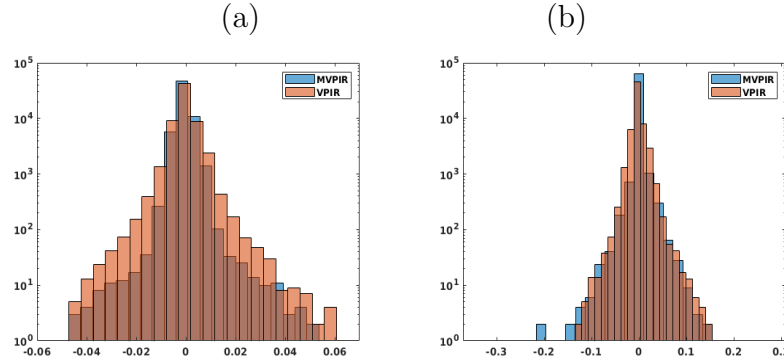
where  $z$  is the observed image,  $\bar{z}$  denotes the mean of  $z$ , and  $\sigma_n$  is the standard deviation of the noise.



**Figure 3.** Sparse MRI experiment, PSNR vs. time plots (in seconds) for VPIR method (red line) and MVPIR method (green line) in the case of  $Nz = 20\%$ .



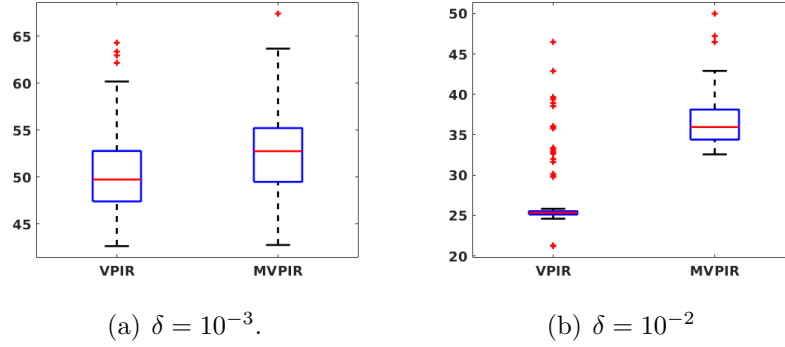
**Figure 4.** Sparse MRI experiment, RMSE vs.  $\lambda$  plots for VPIR method (red line) and MVPIR method (green line) in the case  $Nz = 20\%$ .



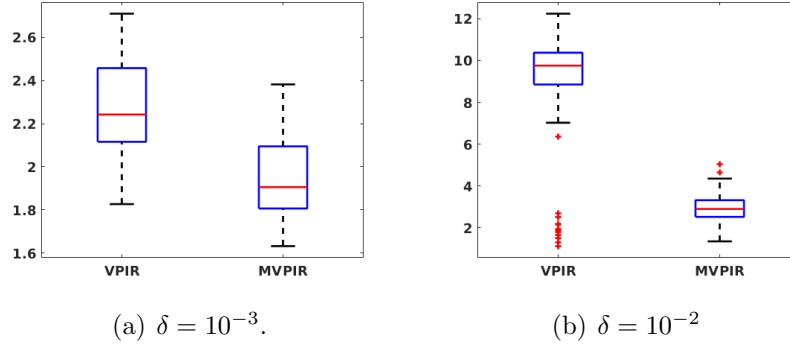
**Figure 5.** Errors histograms in  $\log$  scale, with noise  $\delta = 10^{-2}$  : MVPIR: blue bars. VPIR: pink bars. (a)  $Nz = 20\%$ , (b)  $Nz = 8\%$

The MVPIR loop has been stopped with a discrepancy criterium.

The original QR-code image of size  $256 \times 256$  is shown in Fig. 8(a). We have created the test problem by multiplying the exact image  $U$  by the matrix  $\Phi$  representing the discretization of a Gaussian PSF of size  $13 \times 13$  with standard deviation  $\sigma_g$  ( $\sigma_g = 1.5$  and  $\sigma_g = 3$  in our experiments) and by adding white Gaussian noise of different intensities



**Figure 6.** Boxplot of PSNR obtained by 100 noise realisations for different noise levels:  $\delta = 10^{-3}$  and  $\delta = 10^{-2}$ . Each image reports VPIR on the left and MVPIR on the right. Red line is the median, the top and bottom lines of boxes represent 75 and 25 percentiles, red crosses are outliers.



**Figure 7.** Boxplot of computation time in seconds obtained by 100 noise realisations: for different noise levels:  $\delta = 10^{-3}$  and  $\delta = 10^{-2}$ . Each image reports VPIR on the left and MVPIR on the right. Red line is the median, lines of boxes represent 75 and 25 percentiles, red crosses are outliers.

measured by the BSNR parameter. Two examples of blurred and noisy images are reported in Figures 8(b) and 8(c).

In Table 4 we show the PSNR values of the observed images (column PSNR0) and of the restored images (column PSNR) obtained with the MVPIR and the L2-TV algorithm together with the computation times in seconds, both for noiseless and noisy data. The table confirms the effectiveness of MVPIR method especially when the blur is strong and the noise is high. Figure 9 shows the reconstructions computed by MVPIR and L2-TV algorithms and Figure 10 exhibits a zoom of the previous reconstructions and compare it with the original image. The MVPIR image appears more defined than the L2-TV in the borders of the small squares composing the QR-code, thus confirming the better performance of the nonconvex penalty in this case.

Finally, we report some considerations about the automatic choice of the regularization parameter  $\lambda$  in this deblurring experiment. In Figure 4.2 we plot the PSNR values obtained by the MVPIR method (blu dashed line) and by the L2-TV



$\sigma_g$		PSNR0	MVPIR		L2-TV	
			PSNR	time	PSNR	time
1.5	No Noise	17.26	43.86	29.9	41.98	30.6
	BSNR=30	17.14	31.72	2.22	29.11	1.2
	BSNR=40	17.25	39.78	5.04	33.37	4.7
3	No Noise	13.59	39.48	28.34	33.80	30.7
	BSNR=30	13.55	28.78	10.17	24.71	3.9
	BSNR=40	13.59	34.93	8.26	27.74	3.9

**Table 4.** PSNR and computational times for the deblurring problem with noisy and noiseless data.



(a) Original



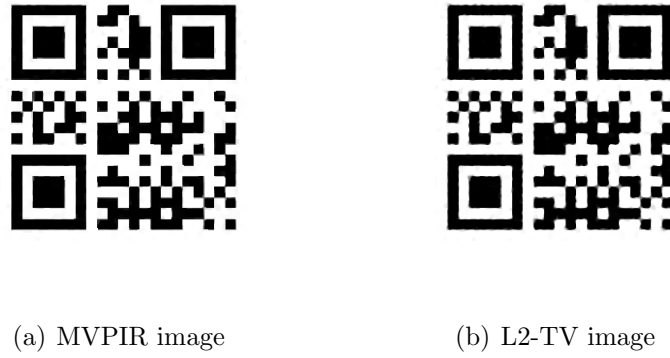
(b) Blurred  $\sigma_g = 1.5, \text{BSNR}=40$ .



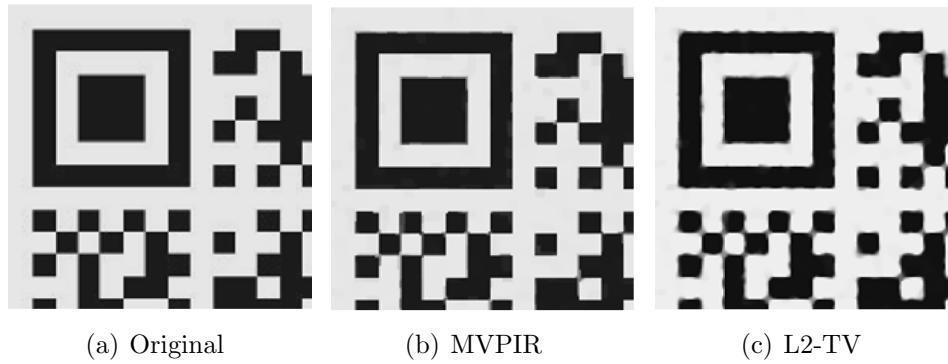
(c) Blurred  $\sigma_g = 3, \text{BSNR}=30$ .

**Figure 8.** Original and blurred QR-code images

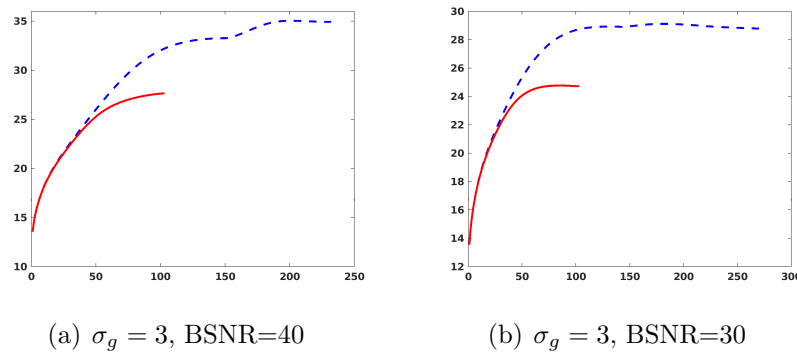
method (red continuous line) as a function of the inner iterations (relative to the Forward Backward solver of (8)) in the case  $\sigma_g = 3$  and BSNR=40 (Figure 11(a)) and BSNR=30 (Figure 11(b)). The non convex penalty exhibits better performance in both cases. In the case of MVPIR the sharp increasing behaviour of the curve means that the rule adopted for the updating of the regularization parameter  $\lambda$  works very well.



**Figure 9.** Deblurred images obtained with IVIPR and L2-TV ( $\sigma_g = 3$ , BSNR=30)



**Figure 10.** Details from the MVPIR and L2-TV reconstructions of Figure 9



**Figure 11.** PSNR values as a function of the iterations for the MVPIR (blue dashed line) and L2-TV (red continuous line) methods.

## 5. Conclusions

In this paper we presented the VPIR algorithm for the solution of a nonconvex penalized minimization problem with the automatic choice of the regularization parameter in imaging applications. We modified the well known IRl1 method by updating the value of the regularization parameter at each iteration and proved its convergence. We presented an implementation where the IRl1 method is repeated for different

nonconvex penalization functions approaching the  $\ell_0$  seminorm. Finally, we considered a more efficient implementation of VPIR. The tests performed on two different imaging applications where the image is sparse in the gradient domain show very good results both in terms of precision and computational efficiency and the choice of the regularization parameter proves to be accurate. In particular, the MVPIR gets the same results of the VPIR but in shorter computational times. The variational formulation considered here is very general and relevant for many imaging applications; moreover, the algorithm is easily extensible to other families of nonconvex regularizers, such as, for example, the  $\ell_p$  norms with  $0 < p < 1$ .

## Acknowledgements

This research has been partially supported by GNCS-INDAM Italy.

## References

- [1] Geman and Geman D 1984 *IEEE Trans. Pattern Anal. Mach. Intell.*, **14** (1992), pp.367-383 **6** 721–741
- [2] Blake A and Zisserman A 1987 *Visual Reconstruction* (MIT Press, Cambridge)
- [3] Mumford D and Shah J 1989 *Comm. Pure Appl. Math.* **42** 577–685
- [4] Chartrand R 2009 Fast algorithms for nonconvex compressive sensing: Mri reconstruction from very few data *IEEE International Symposium on Biomedical Imaging (ISBI)*
- [5] Trzasko J and Manduca A 2009 *IEEE Transactions on Medical Imaging* **28** 106–121
- [6] Sidky E, Chartrand R, Boone J and Pan X 2018 *IEEE Trans. Eng. Health Med.* **2** 1–18
- [7] Candes E, Romberg J and Tao T 2006 *IEEE Tr. Inf. theory* **52** 489–509
- [8] Donoho D 2006 *IEEE Trans. inf. theory* **52** 1289–1306
- [9] Chartrand R 2007 *IEEE Sig. proc. Lett.* **14** 707–710
- [10] Montefusco L B, Lazzaro D and Papi S 2013 *Signal Processing* **93** 2636 – 2647 ISSN 0165-1684
- [11] Wen F, Chu L, Liu P and Qiu R C 2018 *ArXiv e-prints (Preprint 1808.05403)*
- [12] Chartrand R and Yin W 2008 Iteratively reweighted algorithm for compressive sensing *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*
- [13] Ochs P, Dosovitskiy A, Brox T and Pock T 2015 *SIAM Journal on Imaging Sciences* **8** 331–372
- [14] Candes E and Wakin M 2008 *Journal of Fourier Analysis and Applications* **14** 877–905
- [15] Montefusco L B and Lazzaro D 2012 *IEEE Transactions on Image Processing* **21** 1676–1686
- [16] Chambolle A and Pock T 2010 *J. Math. Imag. vision* **40** 120–145
- [17] Goldstein T and Osher S 2009 *SIAM Journal on Imaging Sciences* **2** 323–343
- [18] Wang Y, Yang J, Yin W and Zhang Y 2008 *SIAM Journal on Imaging Sciences* **1** 248–272
- [19] Combettes P L and Wajs V R 2005 *Multiscale Modeling & Simulation* **4** 1168–1200
- [20] Lazzaro D, Piccolomini E L and Zama F 2018 *ArXiv e-prints (Preprint 1809.11135)*
- [21] Lustig M, Donoho D and Pauly J 2007 *Magn. Res. in Medic.* **58** 1182–1198
- [22] Wang S and et al 2016 *Phys. Med. Biol.* **61** 3291–3316
- [23] Shalev-Shwartz S 2011 *Foundations and Trends in Machine Learning* **4** 107–194
- [24] Attouch H, Bolte J and Svaiter B F 2013 *Math. Program Ser. A* **137** 91–129
- [25] Eksioğlu E 2016 *J. Math. Imag. Vis.* **56** 430–440
- [26] Greengard L and Lee J 2004 *SIAM Review* **46** 443–454