A Majorize-Minimize memory gradient method for complex-valued inverse problems

Anisia Florescu, Emilie Chouzenoux, Jean-Christophe Pesquet, Philippe Ciuciu, Silviu Ciochina

A Majorize-Minimize Memory Gradient Method for Complex-Valued Inverse Problems

Anisia Florescu [(a)(1)], Emilie Chouzenoux [(b)(2)], Jean-Christophe Pesquet [(b)], Philippe Ciuciu [(c)] and Silviu Ciochina [(a)] *

June 3, 2013

Abstract

Complex-valued data are encountered in many application areas of signal and image processing. In the context of optimization of functions of real variables, subspace algorithms have recently attracted much interest, owing to their efficiency for solving large-size problems while simultaneously offering theoretical convergence guarantees. The goal of this paper is to show how some of these methods can be successfully extended to the complex case. More precisely, we investigate the properties of the proposed complex-valued Majorize-Minimize Memory Gradient (3MG) algorithm. Important practical applications of these results arise in inverse problems. Here, we focus on image reconstruction in Parallel Magnetic Resonance Imaging (PMRI). The linear operator involved in the observation model then includes a subsampling operator over the k-space (spatial Fourier domain) the choice of which is analyzed through our numerical results. In addition, sensitivity matrices associated with the multiple coil channels come into play. Comparisons with existing optimization methods confirm the good performance of the proposed algorithm.

*(1) Part of A. Florescu's work was supported by PhD Fellowship "Investitii in cercetare-inovare-dezvoltare pentru viitor (DocInvest)", EC project POSDRU/107/1.5/S/76813. (2) Corresponding author. (a) Politehnica University of Bucharest, Telecommunications Dept., Romania. (b) Université Paris-Est, LIGM, UMR CNRS 8049, Champs sur Marne, France. (c) CEA, NeuroSpin center, INRIA Saclay, PARIETAL Team, Gif-sur-Yvette, France.

1 Introduction

Complex-valued data are ubiquitous in signal and image processing. They play a prominent role in applications such as digital communications, spectrum analysis and image recovery. As emphasized in [1], dealing with complex-valued signals raises a number of challenging theoretical issues, in particular owing to their existing relations with the theory of analytic functions. Although it is always possible to transform a complex-valued vector variable in a double-size real-valued vector, this often results in an awkward reformulation of the problem under consideration. This also leads to numerical solutions which may often be less efficient, especially when one makes use of programming languages such as Matlab which are tailored for computations with complex-valued matrices.

Problems involving complex-valued signals are often formulated as the search for a solution satisfying some optimality conditions. Since the related optimization problems do not usually have closed form solutions, efforts have been dedicated to developing specific iterative algorithms for minimizing real-valued functions of complex variables [30, 38]. However, one shortcoming of existing approaches (e.g. interior point methods) is that they may not be very efficient to deal with large-size problems. Another weakness lies in the lack of theoretical convergence guarantees, especially in the nonconvex case. For example, popular methods such as the nonlinear conjugate gradient algorithm, which may be quite effective in practice, have only been proved to converge under restrictive assumptions [34, Chap.5]. Proximal splitting [22] and augmented Lagrangian [3] methods offer more flexibility for minimizing possibly nonsmooth objective functions, but they may suffer from slow convergence. In the case of functions of real variables, a recent majorize-minimize (MM) subspace algorithm has been proposed which overcomes these limitations [17, 18]. This algorithm has been shown to be quite competitive with respect to both continuous and discrete state-of-the-art optimization techniques for solving inverse problems. Moreover, the convergence of the resulting iterates is secured under weak technical assumptions [18]. Note that MM strategies for functions of complex variables were already investigated in [19, 28] but they were restricted to half-quadratic algorithms requiring the inversion of a large-size linear operator, which is not tractable for any acquisition model.

Complex-valued data are involved in several imaging systems such as in Magnetic Resonance Imaging (MRI) which has been at the core of many recent works in the inverse problem community. Most of these works were directed to the proposal of reconstruction methods for parallel MRI (PMRI) [7, 9, 12, 31, 41]. The objective of PMRI is to reduce the acquisition time while maintaining a good image quality. This is achieved by combining subsampling strategies in the k-space with the use of an array of coils so as to compensate spectral decimation with spatial diversity. Let us em-

phasize that the design of an appropriate subsampling scheme is strongly related to compressive sensing issues [10, 32, 36]. Reconstruction approaches based on various variational formulations and optimization algorithms have been proposed for processing PMRI data, most of them being restricted to the convex case. We can mention, in particular, methods based on iterative soft-thresholding [27] or more elaborate proximal algorithms [11, 12], and augmented Lagrangian techniques [2, 7]. Note that, to the best of our knowledge, MM subspace algorithms have never been used in the context of PMRI.

The organization of the paper is as follows: in Section 2.1, the addressed optimization problem is formulated in a general manner and our notation is introduced. In Section 2.2, we recall some classical results about the derivative of real-valued functions of complex variables, which are relevant to this work. Section 2.3 describes the employed MM strategy in the complex case. The proposed complex-valued 3MG (Majorize-Minimize Memory Gradient) algorithm is studied. In particular, its connections with the algorithm in [17, 18] for minimizing functions of real variables are discussed and its convergence properties are analyzed. Section 3 presents the application of our algorithm to PMRI reconstruction by first introducing the multivariate model under consideration, by formulating the optimization problem, and by providing also some details about the algorithm implementation. In Section 3.4, a number of simulation results is provided demonstrating the good performance of the proposed algorithm. Finally, some conclusions are drawn in Section 4.

2 Proposed optimization method

2.1 Problem statement

In this work, we will consider the following penalized optimization problem:

$$\underset{\boldsymbol{x}\in\mathbb{C}^N}{\text{minimize}} \ \big(F(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \Psi(\boldsymbol{x}) + \frac{\varepsilon}{2}\|\boldsymbol{x}\|^2\big), \tag{1}$$

where $\boldsymbol{H} \neq \boldsymbol{0}$ is a matrix in $\mathbb{C}^{Q\times N}$, $\boldsymbol{y}$ is an observation vector in $\mathbb{C}^Q$, $\Phi\colon \mathbb{C}^Q \to \mathbb{R}$, $\Psi\colon \mathbb{C}^N \to \mathbb{R}$, and $\varepsilon \in [0, +\infty)$.

In inverse problems, function Φ usually corresponds to a data-fidelity term and Ψ to a regularization function. The last quadratic term plays a role similar to an elastic net penalization [42].

We focus on the case when function Ψ takes the following form:

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \Psi(\boldsymbol{x}) = \sum_{s=1}^{S} \psi_s(|\boldsymbol{v}_s^{\mathrm{H}}\boldsymbol{x} - c_s|) \tag{2}$$

where $|\cdot|$ denotes the complex modulus, $(\cdot)^{\mathrm{H}}$ is the matrix trans-conjugate operation, and, for every $s \in \{1, \ldots, S\}$, $\psi_s\colon \mathbb{R} \to \mathbb{R}$, $\boldsymbol{v}_s \in \mathbb{C}^N$, and $c_s \in \mathbb{C}$.

Note that (2) models a wide range of regularization functions which are used in practice. When the vectors $(\boldsymbol{v}_s)_{1\leqslant s\leqslant S}$ correspond to local discrete difference operators and $(\psi_s)_{1\leqslant s\leqslant S}$ are identity functions, a complex-valued variant of the anisotropic total variation measure is defined (for images, by computing horizontal and vertical differences, we have thus $S = 2N$). On the other hand, if $(\boldsymbol{v}_s)_{1\leqslant s\leqslant S}$ constitutes a frame of $\mathbb{C}^N$, a so-called frame analysis [25, 29, 35] penalty is obtained. In the special case when $(\boldsymbol{v}_s)_{1\leqslant s\leqslant S}$ is the canonical basis of $\mathbb{C}^N$ $(S = N)$, we get

$$\big(\forall \boldsymbol{x} = (x_n)_{1\leqslant n\leqslant N} \in \mathbb{C}^N\big) \qquad \Psi(\boldsymbol{x}) = \sum_{n=1}^{N} \psi_n(|x_n - c_n|). \tag{3}$$

This kind of Tikhonov regularization [39] is employed to promote the closeness of the target signal to a reference one $\boldsymbol{c} = (c_n)_{1\leqslant n\leqslant N}$.

Notation: For every vector $\boldsymbol{x} \in \mathbb{C}^N$, $\boldsymbol{x}_R \in \mathbb{R}^N$ (resp. $\boldsymbol{x}_I \in \mathbb{R}^N$) denotes the vector of real (resp. imaginary) parts of the components of $\boldsymbol{x}$. Let $\widetilde{\boldsymbol{x}} \in \mathbb{R}^{2N}$ be the "concatenated" vector $\widetilde{\boldsymbol{x}} = [\boldsymbol{x}_R^\top \ \boldsymbol{x}_I^\top]^\top$ where $(\cdot)^\top$ denotes the transpose operation. We define $\widetilde{\Psi}$ the function of real variables associated with Ψ, i.e. $(\forall \boldsymbol{x} \in \mathbb{C}^N)\ \widetilde{\Psi}(\widetilde{\boldsymbol{x}}) = \Psi(\boldsymbol{x})$. A similar notation will be employed for complex-valued matrices and other functions of complex variables. Finally, $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix.

2.2 Complex-valued differential calculus

Let Θ be a function from $\mathbb{C}^N$ to $\mathbb{C}$. According to Wirtinger's calculus [1], the derivative of Θ with respect to the conjugate of its variable is formally defined as

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \nabla\Theta(\boldsymbol{x}) = \frac{1}{2}\Big(\frac{\partial\widetilde{\Theta}(\widetilde{\boldsymbol{x}})}{\partial \boldsymbol{x}_R} + \imath\frac{\partial\widetilde{\Theta}(\widetilde{\boldsymbol{x}})}{\partial \boldsymbol{x}_I}\Big). \tag{4}$$

Throughout this paper, we suppose that:

Assumption 1

(i) $\widetilde{\Phi}$ *is differentiable.*

(ii) *For every* $s \in \{1, \cdots, S\}$, ψ_s *is a differentiable function and* $\lim_{\substack{t\to 0\\ t\neq 0}} \dot{\psi}_s(t)/t \in \mathbb{R}$, *where* $\dot{\psi}_s$ *denotes the derivative of* ψ_s.

The definition in (4) implies that the derivative of $\Phi_{\boldsymbol{H}} = \Phi(\boldsymbol{H}\cdot - \boldsymbol{y})$ is

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \nabla\Phi_{\boldsymbol{H}}(\boldsymbol{x}) = \boldsymbol{H}^{\mathrm{H}}\nabla\Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}). \tag{5}$$

Similarly, it can be deduced from (2) that

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \nabla\Psi(\boldsymbol{x}) = \sum_{s=1}^{S} \dot{\rho}_s(\boldsymbol{v}_s^{\mathrm{H}}\boldsymbol{x} - c_s)\ \boldsymbol{v}_s, \tag{6}$$

where, for every $s \in \{1, \dots, S\}$, $\rho_s = \psi_s(|\cdot|)$, and $\dot{\rho}_s$ is its Wirtinger derivative. For every $s \in \{1, \dots, S\}$ and $z \in \mathbb{C}$, using standard rules of Wirtinger's calculus [26, Table A.1], we have

$$(\forall z \in \mathbb{C}) \qquad \dot{\rho}_s(z) = \frac{1}{2} z \, \omega_s(|z|), \tag{7}$$

where

$$(\forall a \in \mathbb{R}) \qquad \omega_s(a) = \begin{cases} \dfrac{\dot{\psi}_s(a)}{a} & \text{if } a \neq 0 \\ \lim_{\substack{t \to 0 \\ t \neq 0}} \dot{\psi}_s(t)/t & \text{otherwise.} \end{cases} \tag{8}$$

Combined with (6), this shows that

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \quad \nabla \Psi(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{V} \, \mathrm{Diag}\big(\boldsymbol{b}(\boldsymbol{x})\big) \, (\boldsymbol{V}^{\mathrm{H}} \boldsymbol{x} - \boldsymbol{c}) \tag{9}$$

where

$$\boldsymbol{V} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_S] \in \mathbb{C}^{N \times S} \tag{10}$$

and

$$\boldsymbol{b}(\boldsymbol{x}) = (\omega_s(|\boldsymbol{v}_s^{\mathrm{H}} \boldsymbol{x} - c_s|))_{1 \leqslant s \leqslant S}. \tag{11}$$

In summary, the complex-valued derivative of F is

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \nabla F(\boldsymbol{x}) = \boldsymbol{H}^{\mathrm{H}} \nabla \Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}) + \nabla \Psi(\boldsymbol{x}) + \frac{\varepsilon}{2} \boldsymbol{x} \tag{12}$$

where the derivative of Ψ is given by (9).

Remark 1 *In the special case of the Tikhonov regularization function* (3)*,* (9) *reduces to*

$$\big(\forall \boldsymbol{x} = (x_n)_{1 \leqslant n \leqslant N} \in \mathbb{C}^N\big) \quad \nabla \Psi(\boldsymbol{x}) = \frac{1}{2} \big((x_n - c_n) \, \omega_n(|x_n - c_n|)\big)_{1 \leqslant n \leqslant N}. \tag{13}$$

2.3 Quadratic majorization

In order to develop an efficient algorithm for solving Problem (1), we introduce the following additional assumption:

Assumption 2

(i) Φ *has a* β*-Lipschitzian derivative with* $\beta \in (0, +\infty)$*, i.e.*

$$(\forall \boldsymbol{z} \in \mathbb{C}^Q)(\forall \boldsymbol{z}' \in \mathbb{C}^Q) \quad \|\nabla \Phi(\boldsymbol{z}) - \nabla \Phi(\boldsymbol{z}')\| \leqslant \beta \|\boldsymbol{z} - \boldsymbol{z}'\|. \tag{14}$$

(ii) *For every* $s \in \{1, \dots, S\}$*,* $\psi_s(\sqrt{\cdot})$ *is concave on* $[0, +\infty)$*.*

(iii) *There exists* $\overline{\omega} \in [0, +\infty)$ *such that* $(\forall s \in \{1, \dots, S\})$ $(\forall t \in (0, +\infty))$ $0 \leqslant \omega_s(t) \leqslant \overline{\omega}$*, where* ω_s *is defined by* (8)*.*

Note that Assumption 2(i) is quite standard and, in particular, it is satisfied for least squares data fidelity terms. Assumptions 2(ii) and 2(iii) hold for a wide class of penalty functions, e.g. ℓ_2 - ℓ_1 convex functions constituting smooth approximations of the ℓ_1 norm [17], or ℓ_2 -ℓ_0 nonconvex functions providing smooth approximations of the ℓ_0 cost [18].

The following property is the cornerstone of the proposed approach:

Proposition 1 *Under Assumptions 1 and 2, for every* $\boldsymbol{x}' \in \mathbb{C}^N$,

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \quad F(\boldsymbol{x}) \leqslant \Theta(\boldsymbol{x}, \boldsymbol{x}') \tag{15}$$

where

$$\Theta(\boldsymbol{x}, \boldsymbol{x}') = F(\boldsymbol{x}') + 2\,\mathrm{Re}\left\{\nabla F(\boldsymbol{x}')^{\mathrm{H}}(\boldsymbol{x} - \boldsymbol{x}')\right\} + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x}-\boldsymbol{x}') \tag{16}$$

and

$$\boldsymbol{A}(\boldsymbol{x}') = \mu \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \boldsymbol{V}\ \mathrm{Diag}\left(\boldsymbol{b}(\boldsymbol{x}')\right)\boldsymbol{V}^{\mathrm{H}} + \varepsilon \boldsymbol{I}_N \tag{17}$$

with $\mu \in [2\beta, +\infty)$ *(the matrix* $\boldsymbol{V}$ *and the function* $\boldsymbol{b}$ *being defined by* (10) *and* (11)*, respectively).*

The proof of this result is based on properties established for functions of real variables. It is provided in A.

2.4 Subspace algorithm

Subspace algorithms consist of building a sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ according to the following iterative scheme:

$$(\forall k \in \mathbb{N}) \qquad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k, \tag{18}$$

where $\boldsymbol{D}_k \in \mathbb{C}^{N\times M}$ is a subspace search matrix and $\boldsymbol{u}_k \in \mathbb{C}^M$ is a multivariate step-size minimizing $\boldsymbol{u} \mapsto F(\boldsymbol{x}_k + \boldsymbol{D}_k\boldsymbol{u})$ over $\mathbb{C}^M$. The MM strategy replaces the minimization of the original function F over the subspace with successive minimizations of quadratic tangent majorants

$$\boldsymbol{u} \mapsto \Theta(\boldsymbol{x}_k + \boldsymbol{D}_k\boldsymbol{u}, \boldsymbol{x}_k + \boldsymbol{D}_k\boldsymbol{u}') \tag{19}$$

over $\mathbb{C}^M$ for some vector $\boldsymbol{u}' \in \mathbb{C}^M$. The expression of Θ in Proposition 1 shows that, for a given $\boldsymbol{u}'$, an optimal solution is

$$\widehat{\boldsymbol{u}} = \boldsymbol{u}' - 2(\boldsymbol{D}_k^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k\boldsymbol{u}')\boldsymbol{D}_k)^{\dagger}\boldsymbol{D}_k^{\mathrm{H}}\nabla F(\boldsymbol{x}_k + \boldsymbol{D}_k\boldsymbol{u}') \tag{20}$$

where $(\cdot)^{\dagger}$ denotes the pseudo-inverse operation. The resulting complex-valued MM subspace method for solving Problem (1) is given in Algorithm 1.

When the number M of search directions is small, the computation cost of the multivariate step-size $\widehat{\boldsymbol{u}}$ in (20) is reduced since $\boldsymbol{D}_k^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k\boldsymbol{u}')\boldsymbol{D}_k$ is an $M \times M$ matrix. Thus, the complexity of an iteration of the proposed

Algorithm 1 (complex-valued 3MG)

$\boldsymbol{x}_0 \in \mathbb{C}^N$,
For all $k = 0, \ldots$
  $\boldsymbol{u}_k^0 = \boldsymbol{0}$,
  For all $j = 1, \ldots, J$
    $\boldsymbol{B}_k^{j-1} = \boldsymbol{D}_k^{\mathrm{H}} \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^{j-1}) \boldsymbol{D}_k$,
    $\boldsymbol{u}_k^j = \boldsymbol{u}_k^{j-1} - 2(\boldsymbol{B}_k^{j-1})^\dagger \, \boldsymbol{D}_k^{\mathrm{H}} \nabla F(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^{j-1})$,
  $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^J$.

algorithm is quite reasonable. When $M = 2$, a typical choice for the search directions is $\boldsymbol{D}_k = [-\nabla F(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k-1}]$ for every $k \in \mathbb{N}$ (by setting $\boldsymbol{x}_{-1} = \boldsymbol{0}$), which leads to the so-called MM Memory Gradient algorithm.

Algorithm 1 takes a form similar to the one developed in the real case in [17]. Note however that, due to the linear transform property (37), the function defined on $\mathbb{R}^{2M}$ associated with (19) is

$$\widetilde{\boldsymbol{u}} \mapsto \widetilde{\Theta}(\widetilde{\boldsymbol{x}}_k + \widetilde{\boldsymbol{D}}_k \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{x}}_k + \widetilde{\boldsymbol{D}}_k \widetilde{\boldsymbol{u}}') \tag{21}$$

where

$$\widetilde{\boldsymbol{D}}_k = \begin{bmatrix} \boldsymbol{D}_{k,R} & -\boldsymbol{D}_{k,I} \\ \boldsymbol{D}_{k,I} & \boldsymbol{D}_{k,R} \end{bmatrix}. \tag{22}$$

This means that Algorithm 1 can be viewed as a way of expressing in a concise manner [18, Algorithm (3.16)] for minimizing $\widetilde{F}$ over $\mathbb{R}^{2N}$, when the subspace search matrix at iteration k is $\widetilde{\boldsymbol{D}}_k$. Therefore, it corresponds to $2M$ search directions in $\mathbb{R}^{2N}$.

2.5 Convergence analysis

First of all, let us study the existence of a minimizer of function F by making the following assumption:

Assumption 3 *One of the following three conditions holds:*

(i) Φ *and* $(\psi_s)_{1 \leqslant s \leqslant S}$ *are lower bounded functions and* $\varepsilon > 0$.

(ii) (a) Φ *is coercive (i.e.* $\lim_{\|\boldsymbol{z}\| \to +\infty} \Phi(\boldsymbol{z}) = +\infty$*).*
  (b) $(\psi_s)_{1 \leqslant s \leqslant S}$ *are lower bounded functions.*
  (c) $\boldsymbol{H}$ *is injective.*

(iii) (a) Φ *is coercive.*
  (b) *For every* $s \in \{1, \ldots, S\}$, ψ_s *is coercive.*

(c) *The nullspace* $\operatorname{Ker} \boldsymbol{H}$ *of* $\boldsymbol{H}$ *and the orthogonal space* $(\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_S\})^\perp$ *to the space delineated by* $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_S\}$ *are such that*

$$\operatorname{Ker} \boldsymbol{H} \cap (\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_S\})^\perp = \{\boldsymbol{0}\}. \tag{23}$$

Then, we have the following property:

Proposition 2 *Under Assumptions 1 and 3, Problem* (1) *has a solution.*

For the sake of completeness, the proof is given in B although it is based on standard optimization results.

Let us now focus our attention on the convergence properties of the algorithm. A useful tool for analyzing the convergence of iterative optimization algorithms, especially in the nonconvex case, is the Kurdyka-Łojasiewicz inequality [4, 5, 6].

Assumption 4 *F satisfies the Kurdyka-Łojasiewicz inequality i.e. for every $\widehat{\boldsymbol{x}} \in \mathbb{C}^N$ and every bounded neighborhood $\mathbb{B}$ of $\widehat{\boldsymbol{x}}$, there exist three constants $\kappa \in (0, +\infty)$, $\zeta \in (0, +\infty)$ and $\theta \in [0, 1)$ such that*

$$\|\nabla F(\boldsymbol{x})\| \geqslant \kappa |F(\boldsymbol{x}) - F(\widehat{\boldsymbol{x}})|^\theta, \tag{24}$$

for every $\boldsymbol{x} \in \mathbb{B}$ such that $|F(\boldsymbol{x}) - F(\widehat{\boldsymbol{x}})| \leqslant \zeta$ (with the convention $0^0 = 0$).

This inequality is satisfied by many standard functions which are used in practice, in particular semi-algebraic functions. We recall that function F is semi-algebraic if its graph $\{(\boldsymbol{x}, F(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathbb{C}^N\}$ can be expressed as a finite union of subsets defined by a finite number of polynomial inequalities (in the real and imaginary parts of its complex variables). Semi-algebraic functions constitute a wide class of functions. The semi-algebraicity property is stable through standard operations (sum, product, composition,...).

Due to the relation highlighted in Section 2.4 between the complex-valued 3MG algorithm and its real-valued counterpart, the following result can be deduced from [18, Theorem 4.5]:

Proposition 3 *Assume that there exists $\alpha \in (0, +\infty)$ such that $(\forall \boldsymbol{x} \in \mathbb{C}^N)$ $\boldsymbol{A}(\boldsymbol{x}) - \alpha \boldsymbol{I}_N$ is a positive semi-definite matrix. Then, under Assumptions 1-4, Algorithm 1 generates a sequence $(\boldsymbol{x}_k)_{k \in \mathbb{N}}$ converging to a critical point of F. Moreover, $\big(F(\boldsymbol{x}_k)\big)_{k \in \mathbb{N}}$ is a nonincreasing sequence and $(\boldsymbol{x}_k)_{k \in \mathbb{N}}$ has a finite length in the sense that*

$$\sum_{k=0}^{+\infty} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| < +\infty. \tag{25}$$

Some comments can be made about this result:

Remark 2

(i) *We recall that* $\widehat{\boldsymbol{x}}$ *is a critical point of* F *if* $\nabla F(\widehat{\boldsymbol{x}}) = \boldsymbol{0}$. *If* F *is a convex function, Proposition 3 therefore guarantees the convergence of Algorithm 1 to a global minimizer of* F.

(ii) *The positivity assumption on* $\boldsymbol{A}(\boldsymbol{x})$ *for* $\boldsymbol{x} \in \mathbb{C}^N$ *is not restrictive since* (17) *shows that it is satisfied in particular under one of the following conditions:*

  (a) $\varepsilon > 0$.

  (b) $\boldsymbol{H}$ *is injective.*

  (c) *Assumption 3(iii)(c) is satisfied and there exists* $\underline{\omega} \in (0, +\infty)$ *such that* $(\forall s \in \{1, \dots, S\})\ (\forall t \in (0, +\infty))\ \omega_s(t) \geqslant \underline{\omega}$.

It is interesting to note that, in the nonconvex case, the convergence to a global minimizer of F is also guaranteed locally:

Proposition 4 *Suppose that the assumptions of Proposition 3 hold and let* $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ *be a sequence generated by Algorithm 1. There exists* $\eta \in (0, +\infty)$ *such that, if*

$$F(\boldsymbol{x}_0) \leqslant \eta + \inf F, \tag{26}$$

then $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ *converges to a solution to Problem* (1).

The proof of this result is given in C.

3 Application to Parallel MRI

3.1 Model

In parallel MRI, a set of measures $(\boldsymbol{d}_\ell)_{1\leqslant\ell\leqslant L}$ is acquired from L coils. These measures are related to the original full FOV (Field Of View) image $\overline{\boldsymbol{\rho}} \in \mathbb{C}^K$ (the image being columnwise reshaped as a vector) corresponding to a spin density. More precisely, the observation model reads:

$$(\forall \ell \in \{1, \dots, L\}) \qquad \boldsymbol{d}_\ell = \boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{S}_\ell\overline{\boldsymbol{\rho}} + \boldsymbol{w}_\ell \tag{27}$$

where

- for every $\ell \in \{1, \dots, L\}$, $\boldsymbol{S}_\ell \in \mathbb{C}^{K\times K}$ is a diagonal matrix modelling the sensitivity of the coils,
- $\boldsymbol{F} \in \mathbb{C}^{K\times K}$ is a 2D discrete Fourier transform,
- $\boldsymbol{\Sigma} \in \{0,1\}^{\lfloor \frac{K}{R} \rfloor \times K}$ is a subsampling matrix (here, $\lfloor\cdot\rfloor$ designates the rounding operation).

The $\lfloor K/R \rfloor$ lines of matrix $\boldsymbol{\Sigma}$ are thus distinct lines of a $K \times K$ identity matrix, R being the so-called subsampling or acceleration factor.

The noise vectors $(\boldsymbol{w}_\ell)_{1 \leqslant \ell \leqslant L}$ are realizations of random vectors $(\boldsymbol{W}_\ell)_{1 \leqslant \ell \leqslant L}$, which can be assumed mutually statistically independent. In addition, for every $\ell \in \{1, \ldots, L\}$, $\boldsymbol{W}_\ell$ is a circular complex Gaussian vector with zero-mean and covariance matrix $\boldsymbol{\Lambda}_\ell$.

3.2 Variational formulation

In order to provide an estimate of $\overline{\boldsymbol{\rho}}$, we propose to solve the following optimization problem:

$$\underset{\boldsymbol{\rho} \in \mathbb{E}}{\text{minimize}} \quad \sum_{\ell=1}^{L} \|\boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_\ell \boldsymbol{\rho} - \boldsymbol{d}_\ell\|_{\boldsymbol{\Lambda}_\ell^{-1}}^2 + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}} \boldsymbol{\rho}|) \tag{28}$$

where

- $(\forall \ell \in \{1, \ldots, L\})\ \|\cdot\|_{\boldsymbol{\Lambda}_\ell^{-1}}^2 = (\cdot)^{\mathrm{H}} \boldsymbol{\Lambda}_\ell^{-1} (\cdot)$,
- $(\forall s \in \{1, \ldots, S\})\ \psi_s \colon \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{f}_s \in \mathbb{C}^K$,
- $\mathbb{E}$ is a vector subspace corresponding to the range of a matrix $\boldsymbol{E} \in \mathbb{C}^{K \times N}$ with $N \leqslant K$.

The first sum in the above objective function corresponds to the negative-log-likelihood of the noise, while the second one consists of a regularization term. By choosing for $(\boldsymbol{f}_s)_{1 \leqslant s \leqslant S}$ a frame of $\mathbb{C}^K$ (possibly redundant when $S > K$) the above function introduces a so-called frame analysis penalization [25]. The vector space $\mathbb{E}$ serves to incorporate some prior knowledge about the target image. More specifically, we aim at setting to zero the image areas corresponding to the background. This can be performed by choosing $\boldsymbol{E}$ to be an interpolation matrix (i.e. the transpose of a subsampling matrix as defined above) the zero lines of which are associated with pixels belonging to the image background. Such an area can be identified from the sensitivity matrices.

By setting $\boldsymbol{\rho} = \boldsymbol{E}\boldsymbol{x}$ (with $\boldsymbol{x} \in \mathbb{C}^N$) in Problem (28) and adding an elastic net penalization term, the problem can be recast as:

$$\underset{\boldsymbol{x} \in \mathbb{C}^N}{\text{minimize}} \quad \sum_{\ell=1}^{L} \|\boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_\ell \boldsymbol{E} \boldsymbol{x} - \boldsymbol{d}_\ell\|_{\boldsymbol{\Lambda}_\ell^{-1}}^2 + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}} \boldsymbol{E} \boldsymbol{x}|) + \frac{\varepsilon}{2} \|\boldsymbol{x}\|^2, \tag{29}$$

where $\varepsilon \in [0, +\infty)$. It thus appears as an instance of Problem (1) where

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_L \end{bmatrix}, \quad (\forall \ell \in \{1, \ldots, L\}) \;\; \boldsymbol{H}_\ell = \boldsymbol{\Lambda}_\ell^{-1/2} \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_\ell \boldsymbol{E}, \tag{30}$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_L \end{bmatrix}, \qquad (\forall \ell \in \{1, \ldots, L\}) \;\; \boldsymbol{y}_\ell = \boldsymbol{\Lambda}_\ell^{-1/2} \boldsymbol{d}_\ell, \tag{31}$$

$$(\boldsymbol{v}_s)_{1 \leqslant s \leqslant S} = (\boldsymbol{E}^{\mathrm{H}} \boldsymbol{f}_s)_{1 \leqslant s \leqslant S}, \tag{32}$$

$c_s \equiv 0$, and Φ is the squared Hermitian norm of $\mathbb{C}^Q$ with $Q = L\lfloor K/R \rfloor$. Thus, Φ satisfies Assumption 2(i) with $\beta = 1$.

3.3 Algorithm implementation

Due to the form of the minimized criterion, the proposed complex-valued 3MG algorithm is applicable to the reconstruction problem under consideration, provided that functions $(\psi_s)_{1 \leqslant s \leqslant S}$ allow Assumptions 1-4 to be satisfied.

Note that, in order to facilitate the implementation of the algorithm, some recursive relations can be exploited. For simplicity, consider the version of 3MG with $J = 1$ and, for every $k \in \mathbb{N}$, $\boldsymbol{D}_k = [-\nabla F(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k-1}]$.[1] Then, at each iteration $k \in \mathbb{N}$ of the algorithm, by dropping the superscript j, we have

$$\begin{aligned} (\forall \ell \in \{1, \ldots, L\}) \qquad \boldsymbol{o}_{k+1,\ell} &\overset{\Delta}{=} \mu^{1/2} \boldsymbol{H}_\ell (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) \\ &= \mu^{1/2} \boldsymbol{H}_\ell \boldsymbol{D}_k \boldsymbol{u}_k \\ &= [\boldsymbol{m}_{k,\ell}, \boldsymbol{o}_{k,\ell}] \boldsymbol{u}_k, \end{aligned} \tag{33}$$

where $\mu \in [2, +\infty)$ and

$$(\forall \ell \in \{1, \ldots, L\}) \qquad \boldsymbol{m}_{k,\ell} = -\mu^{1/2} \boldsymbol{H}_\ell \nabla F(\boldsymbol{x}_k). \tag{34}$$

This shows that, for every $\ell \in \{1, \ldots, L\}$, vector $\boldsymbol{o}_{k+1,\ell}$ can be computed in a recursive manner, thus allowing to reduce the corresponding computational cost (per iteration) to $2Q$ multiplications instead of NQ. A similar recursion holds for $\boldsymbol{o}_{k+1,L+1} \overset{\Delta}{=} \boldsymbol{V}^{\mathrm{H}} (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$, that is

$$\boldsymbol{o}_{k+1,L+1} = [\boldsymbol{m}_{k,L+1}, \boldsymbol{o}_{k,L+1}] \boldsymbol{u}_k, \tag{35}$$

where $\boldsymbol{m}_{k,L+1} = -\boldsymbol{V}^{\mathrm{H}} \nabla F(\boldsymbol{x}_k)$.

At each iteration k of Algorithm 1, these relations allow us to compute matrix $\boldsymbol{B}_k$ in an efficient manner. This results in the implementation of 3MG in Algorithm 2. Note that each iteration of the algorithm requires the computation of two analysis and one synthesis frame transforms.

[1] Larger values of J were actually observed to lead to slower convergence.

Algorithm 2 (3MG for PMRI reconstruction)

$\boldsymbol{x}_0 \in \mathbb{C}^N, \boldsymbol{x}_{-1} = \boldsymbol{0}$,
For all $k = 0, \ldots$

$$\begin{bmatrix} \boldsymbol{z}_{k,1} \\ \vdots \\ \boldsymbol{z}_{k,L} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_1^{-1/2} \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_1 \\ \vdots \\ \boldsymbol{\Lambda}_L^{-1/2} \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_L \end{bmatrix} \boldsymbol{E} \boldsymbol{x}_k,$$

$$\nabla F(\boldsymbol{x}_k) = \boldsymbol{E}^{\mathrm{H}} \sum_{\ell=1}^{L} \boldsymbol{S}_\ell^{\mathrm{H}} \boldsymbol{F}^{\mathrm{H}} \boldsymbol{\Sigma}^\top \boldsymbol{\Lambda}_\ell^{-1/2} (\boldsymbol{z}_{k,\ell} - \boldsymbol{y}_\ell) + \nabla \Psi(\boldsymbol{x}_k) + \frac{\varepsilon}{2} \boldsymbol{x}_k,$$

$$\begin{bmatrix} \boldsymbol{m}_{k,1} \\ \vdots \\ \boldsymbol{m}_{k,L} \end{bmatrix} = -\mu^{1/2} \begin{bmatrix} \boldsymbol{\Lambda}_1^{-1/2} \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_1 \\ \vdots \\ \boldsymbol{\Lambda}_L^{-1/2} \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_L \end{bmatrix} \boldsymbol{E} \nabla F(\boldsymbol{x}_k),$$

$$\boldsymbol{m}_{k,L+1} = -\boldsymbol{V}^{\mathrm{H}} \nabla F(\boldsymbol{x}_k),$$

If $k = 0$

$$\boldsymbol{o}_{k,\ell} = \mu^{1/2} \boldsymbol{z}_{k,\ell}, \qquad \ell \in \{1, \ldots, L\},$$
$$\boldsymbol{o}_{k,L+1} = \boldsymbol{V}^{\mathrm{H}} \boldsymbol{x}_k,$$

$$\boldsymbol{B}_k = \sum_{\ell=1}^{L} \begin{bmatrix} \boldsymbol{m}_{k,\ell}^{\mathrm{H}} \\ \boldsymbol{o}_{k,\ell}^{\mathrm{H}} \end{bmatrix} [\boldsymbol{m}_{k,\ell}, \boldsymbol{o}_{k,\ell}] + \begin{bmatrix} \boldsymbol{m}_{k,L+1}^{\mathrm{H}} \\ \boldsymbol{o}_{k,L+1}^{\mathrm{H}} \end{bmatrix} \mathrm{Diag}\left(\boldsymbol{b}(\boldsymbol{x}_k)\right) [\boldsymbol{m}_{k,L+1}, \boldsymbol{o}_{k,L+1}] + \varepsilon \boldsymbol{D}_k^{\mathrm{H}} \boldsymbol{D}_k,$$

$$\boldsymbol{u}_k = 2 \boldsymbol{B}_k^\dagger \begin{bmatrix} \|\nabla F(\boldsymbol{x}_k)\|^2 \\ -(\boldsymbol{x}_k - \boldsymbol{x}_{k-1})^{\mathrm{H}} \nabla F(\boldsymbol{x}_k) \end{bmatrix},$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k,$$

$$\boldsymbol{o}_{k+1,\ell} = [\boldsymbol{m}_{k,\ell}, \boldsymbol{o}_{k,\ell}] \boldsymbol{u}_k, \qquad \ell \in \{1, \ldots, L+1\}.$$

3.4 Simulation results

In our experiments, the reconstruction of sagittal views of a 3D anatomical image is performed from noisy parallel MRI data generated according to Model (27). The reference image $\overline{\boldsymbol{\rho}}$ is defined as the reconstruction result from a non-accelerated acquisition ($R = 1$) obtained with a 3 Tesla Siemens Trio magnet having an $L = 32$-channel receiver coil (no parallel transmission has been used). The data have been acquired using a 3D T1-weighted MP-RAGE pulse sequence at a $1 \times 1 \times 1.1$ mm^3 spatial resolution (TE = 2.98 ms, TR = 2300 ms, TI = 900 ms, flip angle = 9°, slice thickness = 1.1 mm, transversal orientation, FOV = $256 \times 240 \times 176$ mm^3, TR between two RF pulses: 7.1 ms, antero-posterior phase encoding). A resampling has been performed in the k-space by zero-filling in order to facilitate the use of fast wavelet decompositions, so leading to a 256×256 image size ($K = 256^2$). Estimates of the sensitivity matrices $(\boldsymbol{S}_\ell)_{1 \leqslant \ell \leqslant L}$ are also available. Fig. 1 illustrates the influence of a subset of these matrices.

Different sampling patterns with $R = 5$ are considered for $\boldsymbol{\Sigma}$, namely regular line subsampling, uniform random, radial, spiral, with π density [15], and polynomial decay of various orders [32] (see Fig. 2). Note that the search for continuous trajectories going through the samples of some distributions has been investigated in some recent works [14, 16, 33].

Finally, a circular complex Gaussian white noise with diagonal covariance matrices $\boldsymbol{\Lambda}_\ell = \sigma^2 \boldsymbol{I}_{\lfloor K/R \rfloor}$, $\ell \in \{1, \ldots, L\}$, is added to the data. The noise variance is here equal to $\sigma^2 = 6 \times 10^9$.

Problem (28) is solved by using Algorithm 2. The convex semi-algebraic ℓ_2 - ℓ_1 penalization function:

$$(\forall t \in \mathbb{R}) \quad \psi_s(t) = \lambda_s(\sqrt{1 + t^2/\delta^2} - 1) \tag{36}$$

is employed, for every $s \in \{1, \cdots, S\}$, with $\lambda_s \in [0, +\infty)$ and $\delta \in (0, +\infty)$. Unless specifically mentioned, in the presented results, $(\boldsymbol{f}_s)_{1 \leqslant s \leqslant S}$ ($S = K$) corresponds to an orthonormal wavelet basis using Symmlet filters of length 10 and the decomposition is performed over 3 resolution levels. For simplicity, the parameters $(\lambda_s)_{1 \leqslant s \leqslant S}$ are equal to the same constant λ for the detail coefficients, while they have been set to zero for the approximation ones. The parameters λ and δ are tuned from one of the slices (No 82) so as to maximize the Signal-to-Noise Ratio (SNR) between the reference image and its reconstructed version ($\lambda = 6 \times 10^{-3}$ and $\delta = 1700$). Fig. 3 illustrates the sensitivity of the approach with respect to the choice of these parameters. For the other slice examples (No 70 and No 121), the values of the parameters determined for the first slice are used.

Table 1 allows us to evaluate the reconstruction performance of 3MG algorithm in terms of SNR for the different sampling patterns. One can observe that sampling strategies based on low-order polynomial distributions

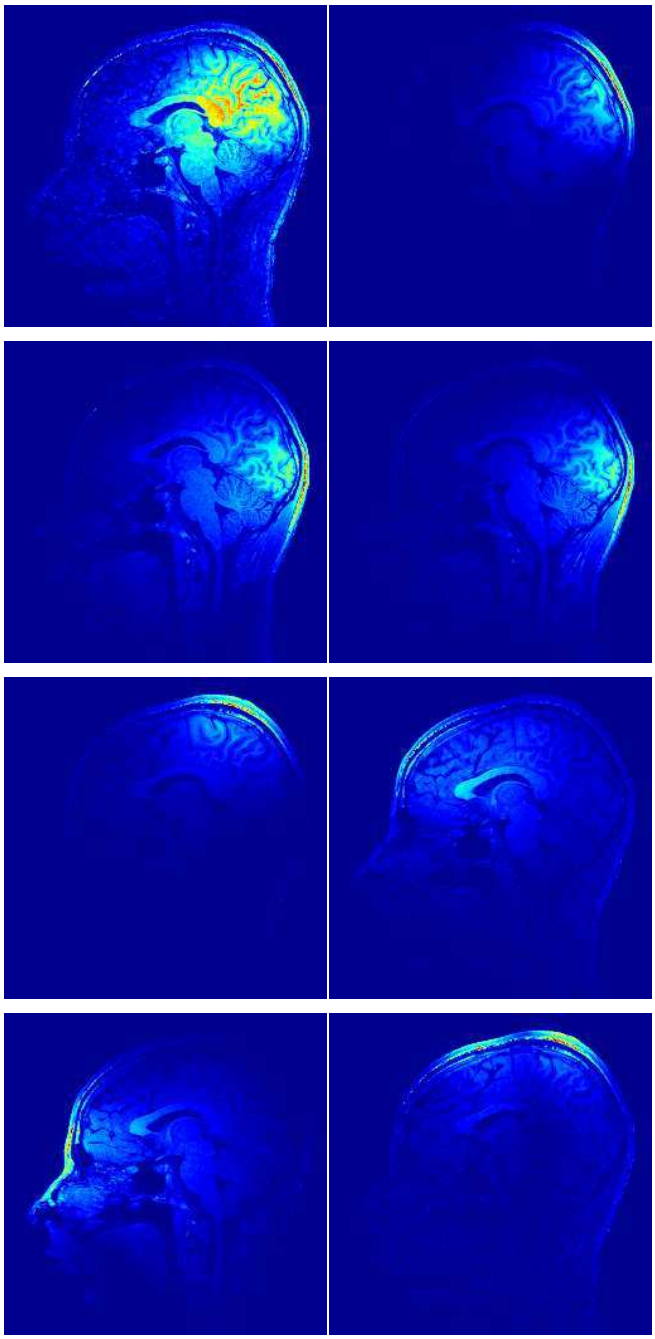

Figure 1: Effects of the sensitivity matrices in the spatial domain in the absence of subsampling ($R = 1$): the moduli of the images corresponding to $(\boldsymbol{S}_\ell \overline{\boldsymbol{\rho}})_{1 \leqslant \ell \leqslant 8}$ are displayed for 8 channels out of 32.

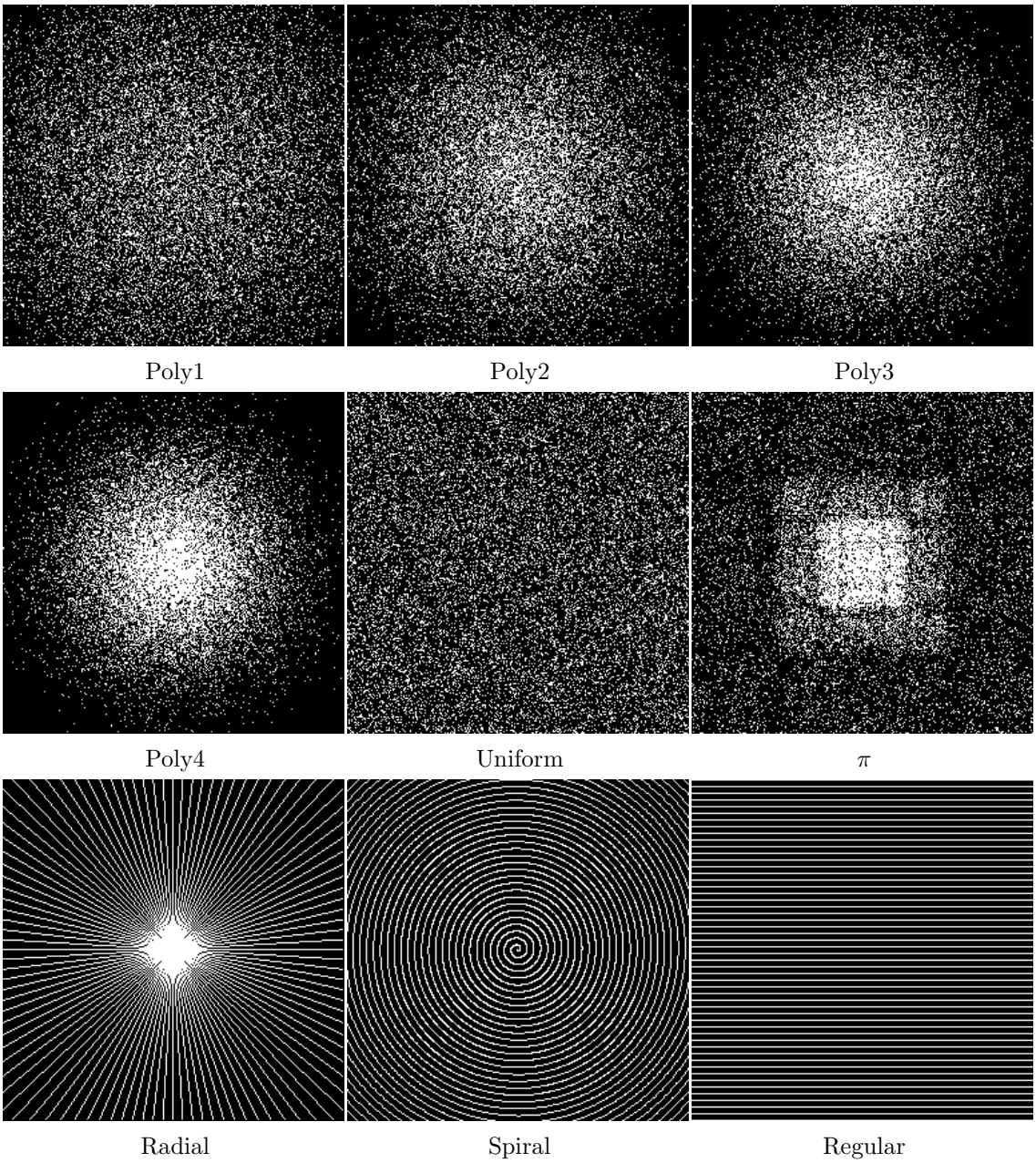

Figure 2: Different types of subsampling.

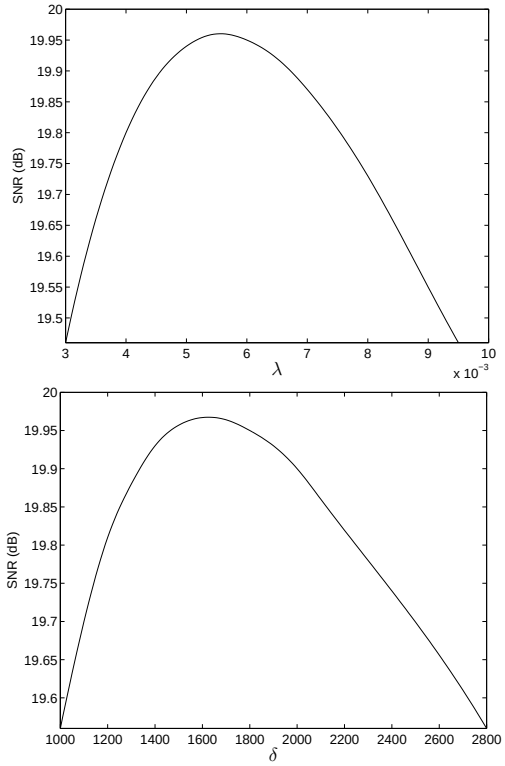

Figure 3: SNR vs λ when $\delta = 1700$ (left) and SNR vs δ when $\lambda = 6 \times 10^{-3}$ (right) for slice No 82 using Poly1 sampling, 3MG algorithm and ℓ_2 - ℓ_1 regularization.

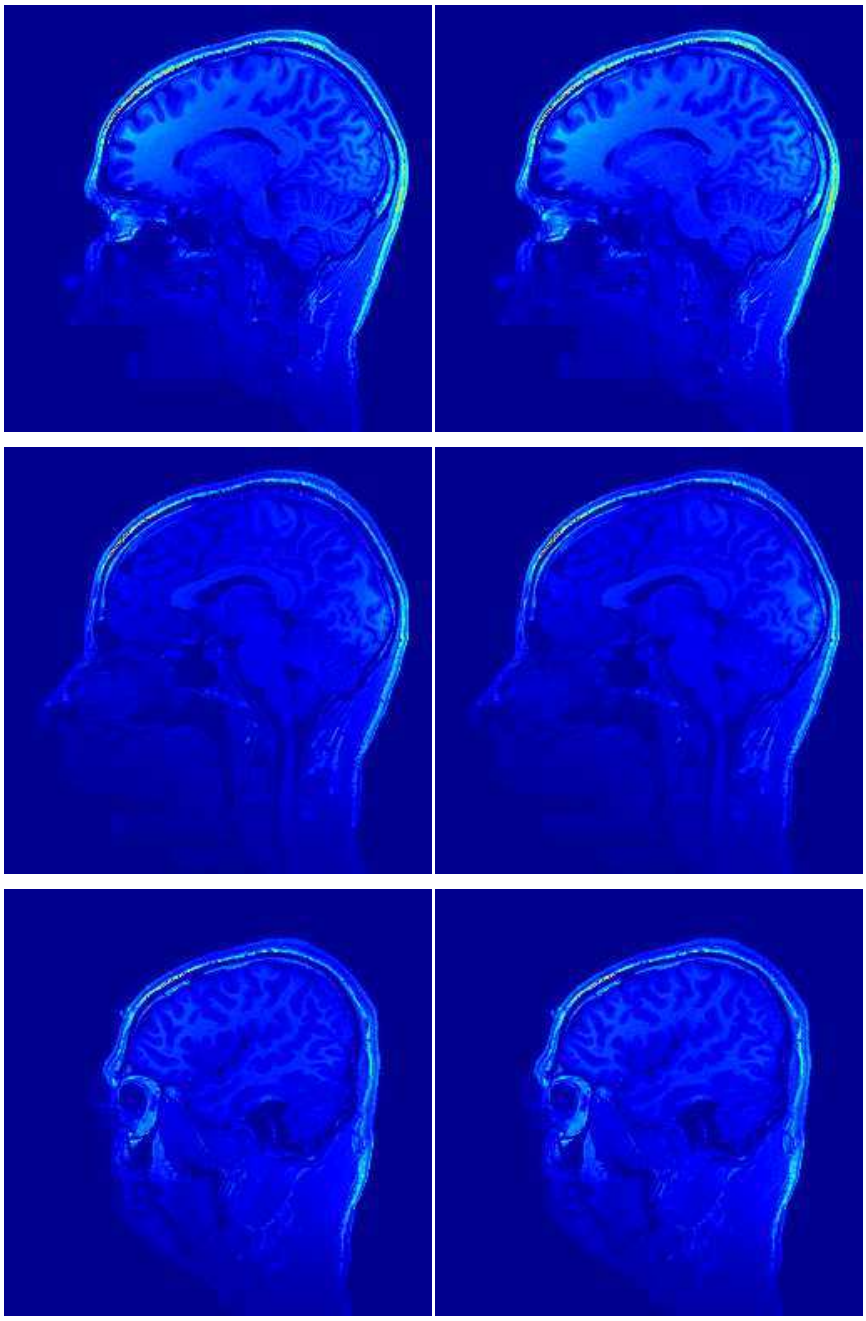

Figure 4: Moduli of the original images $\overline{\boldsymbol{\rho}}$ (left column) and the reconstructed ones (right column) for slices No 70 – SNR = 21.15 dB, No 82 – SNR = 19.96 dB and No 121 – SNR = 20.89 dB (from top to bottom) using Poly1 sampling, 3MG algorithm and ℓ_2 - ℓ_1 regularization.

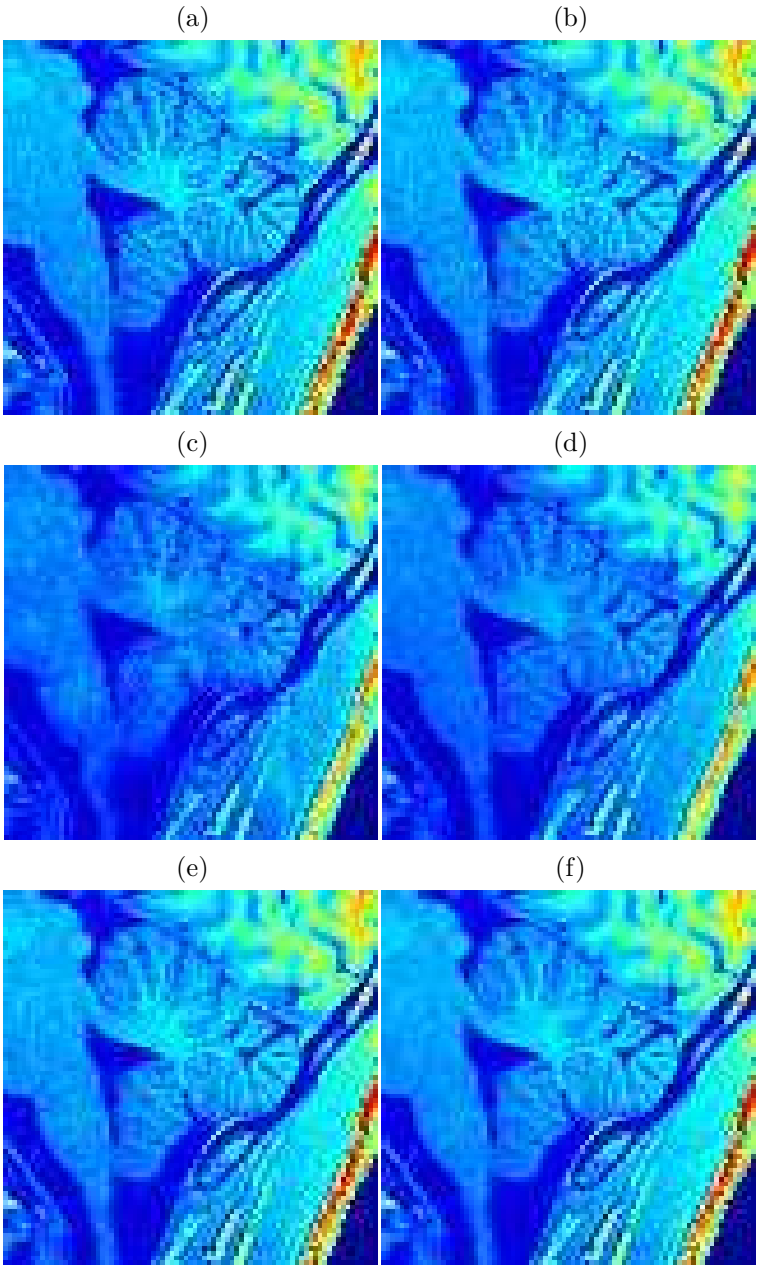

Figure 5: Figs (a)-(f) display zooms centered on the cerebellum area ($[150-230, 100-180]$) for the moduli of the original slice No 82 (a) and reconstructed images using Poly1 (b, e, f), Poly5 (c) or Regular (d) subsampling, 3MG algorithm, ℓ_2 - ℓ_1 regularization (a-d) or ℓ_2 - ℓ_0 (G) regularization (e, f), an othonormal wavelet basis (a-e) or an overcomplete wavelet frame (f).

Table 1: SNR values for various subsampling strategies using 3MG algorithm with ℓ_2 - ℓ_1 regularization.

	SNR (dB)		
Sampling pattern	Slide No 70	Slice No 82	Slice No 121
Poly1	**21.15**	**19.96**	**20.89**
Poly2	20.32	19.34	20.07
Poly3	19.43	18.53	19.18
Poly4	18.47	17.50	18.35
Poly5	17.67	16.95	17.52
Uniform	21.02	19.71	20.68
π	20.46	19.31	20.08
Radial	20.27	19.20	20.01
Spiral	20.35	19.17	20.03
Regular	19.18	18.13	18.66

as well as on the uniform or π distributions lead to higher quality reconstructed images. The moduli of the reconstructed images for Poly1 sampling strategy are displayed in Fig. 4. Visual comparisons between different sampling schemes are also shown in Figs 5 (a)-(d). Note that these results are obtained by setting ε to 0 in (29), since we observed that a nonzero value does not yield a practical improvement of the algorithm performance.

We compare the proposed algorithm with state-of-the-art optimization methods. The first one is the proximal primal-dual method first proposed in [13] and further extended in [24, 40] (here designated by CPCV). The second one is another proximal primal-dual approach, namely the Monotone+Lipschitz Forward-Backward-Forward (M+LFBF) algorithm which was proposed in [23] (see also [8] for extensions). The last competitor is the celebrated Alternating-Direction Method of Multipliers (ADMM) [2, 3, 7]. The matrix inversion involved at each iteration of this algorithm is performed using a few subiterations of a linear conjugate gradient method. All the comparisons are made in terms of computation time for `Matlab R2011b` codes running on a single-core Intel i7-2620M CPU@2.7 GHz with 8 Gb of RAM, in the case of Poly1 sampling strategy.

One of the potential advantages of CPCV, M+LFBF and ADMM is their ability to tackle convex optimization problems involving nonsmooth functions. In our experiments, they are thus employed to solve Problem (29) in the case when an ℓ_1 sparsity promoting regularization is used, that is, for every $s \in \{1, \ldots, S\}$, $\psi_s = \lambda'_s|\cdot|$ with $\lambda'_s \in [0, +\infty)$. Similarly to the ℓ_2 - ℓ_1 penalization, the parameters $(\lambda'_s)_{1 \leqslant s \leqslant S}$ have been chosen equal to the same constant λ' for the detail coefficients, while they have been set to zero for the approximation ones. Although the ℓ_2 - ℓ_1 function can be viewed as a smoothed approximation of the ℓ_1 norm, it is worth noticing that the

Table 2: Reconstruction results for several optimization and regularization strategies using two different decompositions (Poly1 subsampling pattern).

Decomposition	Algorithm	Penalization	SNR (dB)		
			Slice No 70	Slice No 82	Slice No 121
Wavelet basis	M+LFBF	ℓ_1	21.15	19.96	20.89
	CPCV	ℓ_1	21.15	19.96	20.89
	ADMM	ℓ_1	21.15	19.96	20.89
	3MG	ℓ_2 - ℓ_1	21.15	19.96	20.89
	3MG	ℓ_2 - ℓ_0 (H)	21.09	20.05	20.97
	3MG	ℓ_2 - ℓ_0 (W)	21.21	20.17	21.10
	3MG	ℓ_2 - ℓ_0 (G)	21.33	20.27	21.20
Redundant wavelet frame	3MG	ℓ_2 - ℓ_1	21.67	20.46	21.39
	3MG	ℓ_2 - ℓ_0 (G)	**22.10**	**20.94**	**21.84**

resulting SNR values reported in Table 2 are identical for both functions. Moreover, as illustrated by Figs 6 and 7, the proposed algorithm benefits from a faster convergence. Note that in the latter plot showing the convergence profile of the iterates, the asymptotic solution $\widehat{\boldsymbol{x}}$ was precomputed using a large number of iterations.

A further advantage of 3MG is that it allows the use of nonconvex penalizations. We indicate results obtained for some ℓ_2 - ℓ_0 penalizations, namely

- Hyberbolic tangent (H) potential: $(\forall t \in \mathbb{R})\ \psi_s(t) = \lambda_s \tanh\left(\frac{t^2}{2\delta^2}\right)$
- Welsch potential (W): $(\forall t \in \mathbb{R})\ \psi_s(t) = \lambda_s\Big(1 - \exp(-\frac{t^2}{2\delta^2})\Big)$
- Geman-McClure (G) potential: $(\forall t \in \mathbb{R})\ \psi_s(t) = \lambda_s(2\delta^2 + t^2)^{-1}t^2$,

where $\lambda_s \in [0, +\infty)$ and $\delta \in (0, +\infty)$. For these functions, the Kurdyka-Łojasiewicz inequality is satisfied. In addition, it was shown in [18, Proposition 2] that the corresponding global minimizer tends to a minimizer of an ℓ_0 penalized criterion when $\delta \to 0$. In Table 2, it can be noticed a quantitative gain in terms of reconstruction quality with respect to the convex case although the visual improvements are hardly noticeable (see Fig. 5(e)).

Finally, we evaluate the improvements resulting from the use of an undecimated wavelet decomposition carried out over 3 resolutions, still using Symmlet 10 filters. According to our observations, the reconstructed images appear to be a little bit less noisy (see Fig. 5(f)) than when a wavelet basis representation is used. However, it must be pointed out that the reconstruction quality improvement comes at the expense of an increased computational load: the convergence time is about five time larger in our experiments.

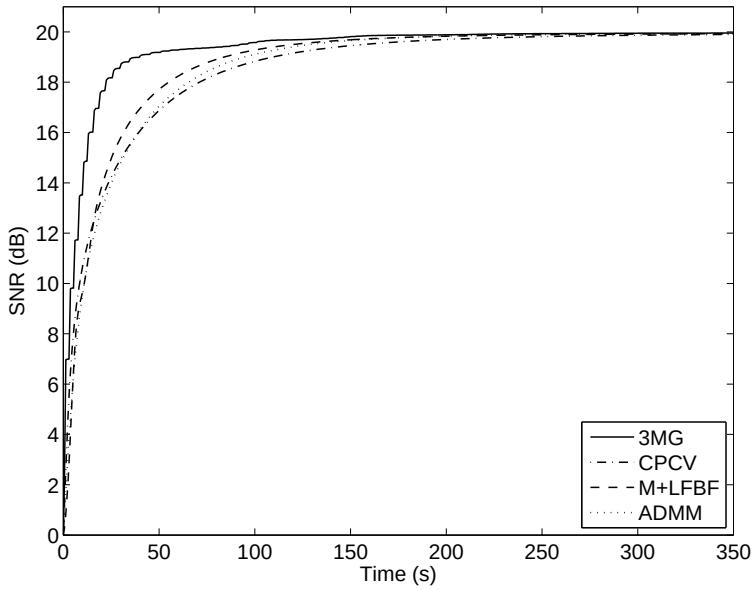

Figure 6: SNR evolution as a function of computation time using 3MG, M+LFBF [23], CPCV [13, 24] and ADMM for slice No 82.

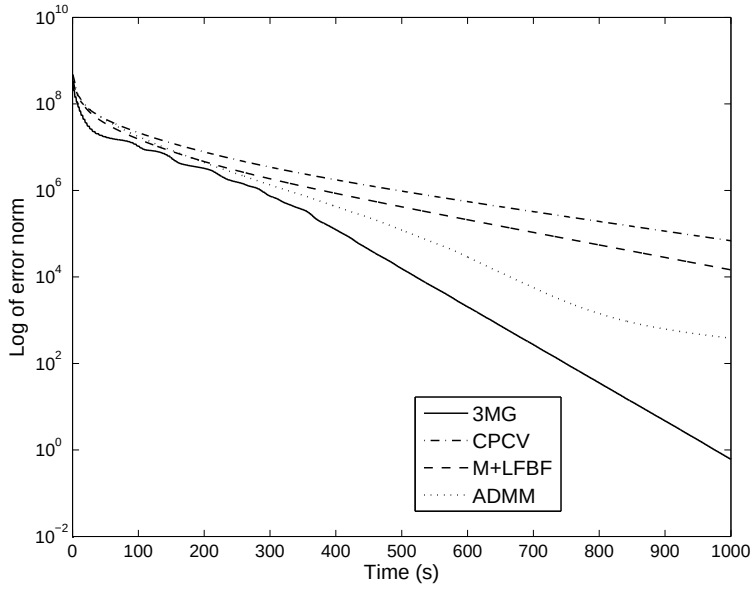

Figure 7: Error $\|\boldsymbol{x}_k - \widehat{\boldsymbol{x}}\|$ (in log scale) as a function of computation time using 3MG, M+LFBF [23], CPCV [13, 24] and ADMM for slice No 82.

4 Conclusion

In this paper, we have proposed an extension of the 3MG algorithm for the solution of large-size optimization problems involving functions of complex variables. We have shown that the proposed algorithm is guaranteed to converge under weak assumptions. Its good numerical performance has been demonstrated in the context of complex-valued image reconstruction from actual parallel MRI data. In particular, comparisons with recent optimization methods have demonstrated the good convergence properties of our algorithm. In addition, we have shown that both convex and nonconvex penalty functions can be considered. We would like to emphasize that the proposed approach makes it possible to employ more sophisticated (non necessarily tight) frame representations [29] than the wavelet decompositions used in our experiments. Moreover, our method can deal with any kind of subsampling strategy. Thus, it can handle sampling schemes currently under development so as to better take into account physical constraints of PMRI systems [14, 16].
Finally, although the focus in this work has been put on PMRI reconstruction, the proposed algorithm is applicable to a range of inverse problems involving complex-valued data, e.g. spectrum analysis problems [20, 21].

A Proof of Proposition 1

Assumption 2(i) is equivalent to say that the function $\widetilde{\Phi}$ of real variables associated with Φ has a (2β)-Lispchitzian gradient. It can be noticed that, for every $\boldsymbol{x} \in \mathbb{C}^N$ and $\boldsymbol{z} \in \mathbb{C}^Q$

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} \quad \Leftrightarrow \quad \widetilde{\boldsymbol{z}} = \widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{x}} \tag{37}$$

where

$$\widetilde{\boldsymbol{H}} = \begin{bmatrix} \boldsymbol{H}_R & -\boldsymbol{H}_I \\ \boldsymbol{H}_I & \boldsymbol{H}_R \end{bmatrix}. \tag{38}$$

We recall here that $\boldsymbol{H}_R \in \mathbb{R}^{Q\times N}$ (resp. $\boldsymbol{H}_I \in \mathbb{R}^{Q\times N}$) denotes the real (resp. imaginary) part of the matrix $\boldsymbol{H}$. We have thus

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}) = \widetilde{\Phi}(\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{y}}). \tag{39}$$

In addition, for every $s \in \{1, \ldots, S\}$ and $\boldsymbol{x} \in \mathbb{C}^N$,

$$\psi_s(|\boldsymbol{v}_s^{\mathrm{H}}\boldsymbol{x} - c_s|) = \psi_s(\|\widetilde{\boldsymbol{V}}_s\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{c}}_s\|) \tag{40}$$

where

$$\widetilde{\boldsymbol{V}}_s = \begin{bmatrix} \boldsymbol{v}_{s,R}^\top & \boldsymbol{v}_{s,I}^\top \\ -\boldsymbol{v}_{s,I}^\top & \boldsymbol{v}_{s,R}^\top \end{bmatrix} \tag{41}$$

and $\widetilde{\boldsymbol{c}}_s = [c_{s,R}, c_{s,I}]^\top$. We deduce then from [18, Lemma 1] that, for every $\boldsymbol{x}' = (x'_n)_{1\leqslant n\leqslant N} \in \mathbb{C}^N$,

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad F(\boldsymbol{x}) = \widetilde{F}(\widetilde{\boldsymbol{x}}) \leqslant \widetilde{\Theta}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{x}}') \tag{42}$$

where

$$\widetilde{\Theta}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{x}}') = \widetilde{F}(\widetilde{\boldsymbol{x}}') + \nabla\widetilde{F}(\widetilde{\boldsymbol{x}}')^\top(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') + \frac{1}{2}(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}')^\top \widetilde{\boldsymbol{A}}(\widetilde{\boldsymbol{x}}')(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') \tag{43}$$

$$\widetilde{A}(\widetilde{\boldsymbol{x}}') = \mu\widetilde{\boldsymbol{H}}^\top\widetilde{\boldsymbol{H}} + \sum_{s=1}^{S} b_s(\boldsymbol{x}')\widetilde{\boldsymbol{V}}_s^\top\widetilde{\boldsymbol{V}}_s + \varepsilon\boldsymbol{I}_{2N} \tag{44}$$

and, for every $s \in \{1, \ldots, S\}$, $b_s(\boldsymbol{x}') = \omega_s(\|\widetilde{\boldsymbol{V}}_s\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{c}}_s\|) = \omega_s(|\boldsymbol{v}_s^{\mathrm{H}}\boldsymbol{x} - c_s|)$. We have then

$$\begin{aligned}\widetilde{\Theta}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{x}}') = \widetilde{F}(\widetilde{\boldsymbol{x}}') + \nabla\widetilde{F}(\widetilde{\boldsymbol{x}}')^\top(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') + \mu\big(\widetilde{\boldsymbol{H}}(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}')\big)^\top\widetilde{\boldsymbol{H}}(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') \\ + \frac{1}{2}\sum_{s=1}^{S} b_s(\boldsymbol{x}')\left(\widetilde{\boldsymbol{V}}_s(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}')\right)^\top\widetilde{\boldsymbol{V}}_s(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') + \frac{\varepsilon}{2}\|\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}'\|^2.\end{aligned} \tag{45}$$

By using the fact that

$$\big(\forall(\boldsymbol{a}, \boldsymbol{a}') \in (\mathbb{C}^N)^2\big) \qquad \mathrm{Re}\{\boldsymbol{a}^{\mathrm{H}}\boldsymbol{a}'\} = \widetilde{\boldsymbol{a}}^\top\widetilde{\boldsymbol{a}}', \tag{46}$$

and by employing the definition of the Wirtinger derivative in (4) and the linear transform property in (37)-(38), we get

$$\nabla\widetilde{F}(\widetilde{\boldsymbol{x}}')^\top(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') = 2\mathrm{Re}\{\nabla F(\boldsymbol{x}')^{\mathrm{H}}(\boldsymbol{x} - \boldsymbol{x}')\} \tag{47}$$

$$\begin{aligned}\big(\widetilde{\boldsymbol{H}}(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}')\big)^\top\widetilde{\boldsymbol{H}}(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') &= \mathrm{Re}\left\{\big(\boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}')\big)^{\mathrm{H}}\boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}')\right\} \\ &= (\boldsymbol{x} - \boldsymbol{x}')^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}')\end{aligned} \tag{48}$$

and

$$\begin{aligned}&\sum_{s=1}^{S} b_s(\boldsymbol{x}')\left(\widetilde{\boldsymbol{V}}_s(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}')\right)^\top\widetilde{\boldsymbol{V}}_s(\widetilde{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}') \\ &= \sum_{s=1}^{S} b_s(\boldsymbol{x}')\left(\boldsymbol{v}_s^{\mathrm{H}}(\boldsymbol{x} - \boldsymbol{x}')\right)^*\boldsymbol{v}_s^{\mathrm{H}}(\boldsymbol{x} - \boldsymbol{x}') \\ &= (\boldsymbol{x} - \boldsymbol{x}')^{\mathrm{H}}\Big(\sum_{s=1}^{S} b_s(\boldsymbol{x}')\boldsymbol{v}_s\boldsymbol{v}_s^{\mathrm{H}}\Big)(\boldsymbol{x} - \boldsymbol{x}') \\ &= (\boldsymbol{x} - \boldsymbol{x}')^{\mathrm{H}}\boldsymbol{V}\ \mathrm{Diag}\left(\boldsymbol{b}(\boldsymbol{x}')\right)\boldsymbol{V}^{\mathrm{H}}(\boldsymbol{x} - \boldsymbol{x}').\end{aligned} \tag{49}$$

The majorization in (15) and (16) readily follows.

B Proof of Proposition 2

Proof: Under Assumptions 3(iii)(a) and 3(iii)(b),

$$(\boldsymbol{z},\boldsymbol{t}) = \big(\boldsymbol{z},(t_s)_{1\leqslant s\leqslant S}\big) \mapsto \Phi(\boldsymbol{z}-\boldsymbol{y}) + \sum_{s=1}^{S} \psi(|t_s - c_s|)$$

is coercive. Assumption 3(iii)(c) means that $\operatorname{Ker}\boldsymbol{H} \cap \operatorname{Ker}(\boldsymbol{V}^{\mathrm{H}}) = \{\boldsymbol{0}\}$, which is equivalent to the fact that $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sum_{s=1}^{S} \boldsymbol{v}_s\boldsymbol{v}_s^{\mathrm{H}} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}$ is invertible. Consequently, if $\|\boldsymbol{x}\| \to +\infty$ and

$$\begin{bmatrix}\boldsymbol{z}\\ \boldsymbol{t}\end{bmatrix} = \begin{bmatrix}\boldsymbol{H}\\ \boldsymbol{V}^{\mathrm{H}}\end{bmatrix}\boldsymbol{x} \tag{50}$$

then

$$\|\boldsymbol{z}\|^2 + \|\boldsymbol{t}\|^2 = \boldsymbol{x}^{\mathrm{H}}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}})\boldsymbol{x} \geqslant \lambda_{\min}\|\boldsymbol{x}\|^2 \to +\infty, \tag{51}$$

where $\lambda_{\min} > 0$ is the smallest eigenvalue of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \boldsymbol{V}\boldsymbol{V}^{\mathrm{H}}$. This allows us to deduce that F is coercive. In addition, according to Assumption 1, F is a continuous function. The existence of a minimizer then follows from [37, Theorem 1.9].

The same conclusion holds under Assumption 3(i) or 3(ii) which also ensures the coercivity of F.

C Proof of Proposition 4

According to Proposition 2, there exists $\widehat{\boldsymbol{x}} \in \mathbb{C}^N$ such that $F(\widehat{\boldsymbol{x}}) = \inf F$. Furthermore, we have shown in B that F is coercive. This implies that, for every $\upsilon \in (0,+\infty)$,

$$\mathbb{B} = \{\boldsymbol{x} \in \mathbb{C}^N \mid F(\boldsymbol{x}) \leqslant F(\widehat{\boldsymbol{x}}) + \upsilon\} \tag{52}$$

is a bounded set. Thus, according to Assumption 4, there exist $\kappa \in (0,+\infty)$, $\zeta \in (0,+\infty)$ and $\theta \in [0,1)$ such that (24) holds for every $\boldsymbol{x} \in \mathbb{B}$ such that $|F(\boldsymbol{x}) - F(\widehat{\boldsymbol{x}})| \leqslant \zeta$. Since $\widehat{\boldsymbol{x}}$ is a global minimizer of F, the latter inequality is equivalent to $F(\boldsymbol{x}) \leqslant F(\widehat{\boldsymbol{x}}) + \zeta$. Let us now set $\eta = \min\{\upsilon,\zeta\}$.

According to Proposition 2, $\big(F(\boldsymbol{x}_k)\big)_{k\in\mathbb{N}}$ is a nondecreasing sequence. Then, if (26) is satisfied, $(\forall k \in \mathbb{N})$ $F(\boldsymbol{x}_k) \leqslant F(\widehat{\boldsymbol{x}}) + \eta$ and

$$\|\nabla F(\boldsymbol{x}_k)\| \geqslant \kappa |F(\boldsymbol{x}_k) - F(\widehat{\boldsymbol{x}})|^{\theta}. \tag{53}$$

In addition, Proposition 2 allows us to claim that $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ converges to a critical point $\widehat{\boldsymbol{x}}'$ of F. It can be deduced from the continuity of F that $F(\boldsymbol{x}_k) \to F(\widehat{\boldsymbol{x}}')$ and the closedness property of the gradient implies that $\nabla F(\boldsymbol{x}_k) \to \nabla F(\widehat{\boldsymbol{x}}') = \boldsymbol{0}$. Hence, by taking the limit in (53), we get

$$\kappa |F(\widehat{\boldsymbol{x}}') - F(\widehat{\boldsymbol{x}})|^{\theta} = 0 \tag{54}$$

which shows us that $F(\widehat{\boldsymbol{x}}') = F(\widehat{\boldsymbol{x}})$, that is the limit $\widehat{\boldsymbol{x}}'$ of $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ is a global minimizer of F.

Acknowledgement

The authors would like to thank N. Chauffert and S. Mériaux from NeuroSpin (CEA) for their help in providing the MRI data and the subsampling schemes.

References

[1] T. Adali, P. J. Schreier, and L. L. Scharf. Complex-valued signal processing: the proper way to deal with impropriety. *IEEE Trans. Signal Process.*, 59(11):5101–5125, Nov. 2011.

[2] J. Aelterman, H. Q. Luong, B. Goossens, A. Pizurica, and W. Philips. Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data. *Signal Process., Special Issue: Advances in Multirate Filter Bank Structures and Multiscale Representations*, 91(12):2731–2742, Dec. 2011.

[3] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian aproach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20(3):681–695, 2011.

[4] H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Prog.*, 116(1-2):5–16, Jan. 2009.

[5] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, 2010.

[6] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Prog.*, 137(1-2):1–39, Feb. 2013.

[7] C. Bilen, Y. Wang, and I. W. Selesnick. High-speed compressed sensing reconstruction in dynamic parallel MRI using augmented Lagrangian and parallel processing. *IEEE J. Emerging Sel. Top. Circuits Syst.*, 2(3):370–379, Sep. 2012.

[8] R. I Boţ and C. Hendrich. Convergence analysis for a primal-dual monotone + skew splitting algorithm with application to total variation minimization. Preprint Univ. Chemnitz, `http://arxiv.org/abs/1211.1706`, 2012.

[9] C. Boyer, P. Ciuciu, P. Weiss, and S. Mériaux. HYR2PICS: hybrid regularized reconstruction for combined parallel imaging and compressive sensing in MRI. In *Proc. 9th IEEE Int. Symp. Biomed. Imaging (ISBI'12)*, pages 66–69, 2-5 May 2012.

[10] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489–509, 2006.

[11] L. Chaâri, S. Mériaux, S. Badillo, P. Ciuciu, and J.-C. Pesquet. 3D wavelet-based regularization for parallel MRI reconstruction: impact on subject and group-level statistical sensitivity in fMRI. In *Proc. 8th IEEE Int. Symp. Biomed. Imaging (ISBI'11)*, pages 460–464, Chicago, IL, 30 Mar.- 2 Apr. 2011.

[12] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu. A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging. *Med. Image Anal.*, 15(2):185–201, Nov. 2011.

[13] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2010.

[14] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss. Traveling salesman-based variable density sampling. In *Proc. 10th Int. Conf. Sampl. Theory App. (SampTA'13)*, Bremen, Germany, 1-5 Jul. 2013.

[15] N. Chauffert, P. Ciuciu, and P. Weiss. Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies. In *Proc. 10th IEEE Int. Symp. Biomed. Imaging (ISBI'13)*, 7-11 Apr. 2013.

[16] N. Chauffert, P. Ciuciu, P. Weiss, and F. Gamboa. From variable density sampling to continuous sampling using Markov chains. In *Proc. 10th Int. Conf. Sampl. Theory App. (SampTA'13)*, Bremen, Germany, 1-5 Jul. 2013.

[17] E. Chouzenoux, J. Idier, and S. Moussaoui. A majorize-minimize subspace strategy for subspace optimization applied to image restoration. *IEEE Trans. Image Process.*, 20(18):1517–1528, Jun. 2011.

[18] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization. *SIAM J. Imag. Sci.*, 6(1):563–591, 2013.

[19] P. Ciuciu and J. Idier. A half-quadratic blockcoordinate descent method for spectral estimation. *Signal Process.*, 82(7):941–959, 2002.

[20] P. Ciuciu and J. Idier. Regularized Doppler radar imaging for target identification in atmospheric clutter. In *proc. IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP'04)*, volume 5, Montreal, Canada, 17-21 May 2004.

[21] P. Ciuciu, J. Idier, and J.-F. Giovannelli. Regularized estimation of mixed spectra using a circular Gibbs-Markov model. *IEEE Trans. Signal Process.*, 49(10):2202–2213, Oct. 2001.

[22] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2010.

[23] P. L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.*, 20(2):307–330, Jun. 2012.

[24] L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *to appear in J. Optim. Theory Appl*, Dec. 2012. `hal.archives-ouvertes.fr/hal-00609728`.

[25] M. Elad, P. Milanfar, and R. Ron. Analysis versus synthesis in signal priors. *Inverse Prob.*, 23(3):947–968, Jun. 2007.

[26] R. F. H. Fischer. *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, Inc., 2002.

[27] M. Guerquin-Kern, M. Häberlin, K. P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 30(9):1649–1660, Sep. 2011.

[28] S. Husse, Y. Goussard, and J. Idier. Extended forms of Geman Yang algorithm: application to MRI reconstruction. In *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP'04)*, volume 3, pages 513–516, 17-21 May 2004.

[29] L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12):26992730, 2011. Special issue on Advances in Multirate Filter Bank Structures and Multiscale Representations.

[30] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE J. Sel. Top. Sign. Proces.*, 1(4):606–617, Dec. 2007.

[31] B. Liu, F. M. Sebert, Y. Zou, and L. Ying. SparseSENSE: randomly-sampled parallel imaging using compressed sensing. In *Proc. 16th Annual Meeting of ISMRM*, page 3154, 3-8 May 2008.

[32] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58:1182–1195, 2007.

[33] M. Lustig, S. J. Kim, and J. M. Pauly. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *IEEE Trans. Med. Imag.*, 27(6):866–873, 2008.

[34] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, NY, 1999.

[35] N. Pustelnik, J.-C. Pesquet, and C. Chaux. Relaxing tight frame condition in parallel proximal methods for signal restoration. *IEEE Trans. on Signal Proc.*, 60(2):968 – 973, Feb. 2012.

[36] G. Puy, J. P. Marques, R. Gruetter, J. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread spectrum magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 31(3):586–598, Mar. 2012.

[37] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, 1st edition, 1997.

[38] L. Sorber, M. Van Barel, and L. De Lathauwer. Unconstrained optimization of real functions in complex variables. *SIAM J. Optim.*, 22(3):879–898, Jul. 2012.

[39] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston, Washington, DC, 1977.

[40] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, Apr. 2013.

[41] X. Ye, Y. Chen, W. Lin, and F. Huang. Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories. *IEEE Trans. Med. Imag.*, 30(3):575–585, 2011.

[42] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*, 67(2):301–320, 2005.