A FORWARD-BACKWARD ALGORITHM FOR REWEIGHTED PROCEDURES: APPLICATION TO RADIO-ASTRONOMICAL IMAGING

Audrey Repetti[⋆†]*, and Yves Wiaux*[⋆⋆]

[⋆] Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom
[†] Department of Actuarial Mathematics & Statistics, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

ABSTRACT

During the last decades, reweighted procedures have shown high efficiency in computational imaging. They aim to handle non-convex composite penalization functions by iteratively solving multiple approximated sub-problems. Although the asymptotic behaviour of these methods has recently been investigated in several works, they all necessitate the sub-problems to be solved accurately, which can be sub-optimal in practice. In this work we present a reweighted forward-backward algorithm designed to handle non-convex composite functions. Unlike existing convergence studies in the literature, the weighting procedure is directly included within the iterations, avoiding the need for solving any sub-problem. We show that the obtained reweighted forward-backward algorithm converges to a critical point of the initial objective function. We illustrate the good behaviour of the proposed approach on a Fourier imaging example borrowed to radio-astronomical imaging.

Index Terms— Non-convex optimization, forward-backward algorithm, reweighted procedure, Fourier imaging, astronomical imaging.

1. INTRODUCTION

In the context of inverse problems, a common and yet efficient approach to find an estimate $\boldsymbol{x}^\dagger$ of the unknown object of interest $\overline{\boldsymbol{x}} \in \mathbb{R}^N$, is to define it as a minimizer of a sum of two functions: the data-fidelity term F associated with the statistical model, and the regularization term R incorporating *a priori* information on $\overline{\boldsymbol{x}}$:

$$\boldsymbol{x}^\dagger \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}} \ \{H(\boldsymbol{x}) = F(\boldsymbol{x}) + R(\boldsymbol{x})\}. \qquad (1)$$

We focus on the case where $F\colon \mathbb{R}^N \to \mathbb{R}$ is Lipschitz-differentiable, and $R\colon \mathbb{R}^N \to]-\infty, +\infty]$ is a non-smooth function that can be written as a sum of compositions of two functions. This type of composite function R usually leads to reweighting procedures, widely used in signal and image processing [1–5]. They consist in promoting sparsity by mimicking the ℓ_0 pseudo-norm without solving explicitly any non-convex problem. To this aim, they successively solve penalized minimization problems, changing each time the weights appearing in the penalization function. As a well-known example, we can mention the reweighted ℓ_1 minimization approach [6], aiming to minimize a log-sum prior by successively solving weighted-ℓ_1 sub-problems. The convergence of reweighted methods have been investigated during the last decades, e.g., in [7–16]. Although all these works investigate the convergence of the overall methods under different technical assumptions, in general they all rely on the same strategy, consisting in iteratively minimizing approximations to the objective function. The common main issue of these methods is that they necessitate to iteratively solve accurately multiple sub-problems, using iterative algorithms.

In this work, we directly incorporate the reweighting process into a unique iterative algorithm, namely the forward-backward algorithm [17–20], avoiding the need for solving accurately sub-problems. More precisely, the proposed method, deduced from our previous work [21], alternates at each iteration between a gradient step on F and a proximity step handling an approximation to the penalization function R. The approximation of R can be updated at each iteration, or kept unchanged for multiple successive iterations. The resulting algorithm is ensured to converge to a critical point of H. We will show empirically, through simulations on a radio-astronomical imaging problem, that the flexibility of choosing the accuracy of solving the sub-problems leads to both, more accurate results, and an acceleration of the convergence of the overall reweighted procedure, compared with traditional reweighted procedures described in the above paragraphs.

The remainder of this work is organized as follows: In Section 2 we present the optimization problem of interest and the proposed algorithm. Simulation results are given in Section 3. Finally, we conclude in Section 4.

[⋆]This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grants EP/M011089/1, EP/M008843/1 and EP/M019306/1).

Penalization	$\psi_p(\boldsymbol{x})$	$\phi_p(u)$	$R(\boldsymbol{x})$	$\lambda_{p,k}$
Log-sum	$\lvert[\boldsymbol{W}\boldsymbol{x}]^{(p)}\rvert$	$\log(u+\varepsilon)$	$\sum_{p=1}^{P}\lambda_p \log\big(\lvert[\boldsymbol{W}\boldsymbol{x}]^{(p)}\rvert+\varepsilon\big)$	$\lambda_p\big(\lvert[\boldsymbol{W}\boldsymbol{x}_k]^{(p)}\rvert+\varepsilon\big)^{-1}$
Cauchy	$\big([\boldsymbol{W}\boldsymbol{x}]^{(p)}\big)^2$	$\log(u+\varepsilon)$	$\sum_{p=1}^{P}\lambda_p \log\Big(\big([\boldsymbol{W}\boldsymbol{x}]^{(p)}\big)^2+\varepsilon\Big)$	$\lambda_p\Big(\big([\boldsymbol{W}\boldsymbol{x}_k]^{(p)}\big)^2+\varepsilon\Big)^{-1}$
Semi-smoothed ℓ_ρ-norm, $0<\rho<1$	$\lvert[\boldsymbol{W}\boldsymbol{x}]^{(p)}\rvert$	$(u+\varepsilon)^\rho-\varepsilon^\rho$	$\sum_{p=1}^{P}\lambda_p\big(\lvert[\boldsymbol{W}\boldsymbol{x}]^{(p)}\rvert+\varepsilon\big)^\rho-\varepsilon^\rho$	$\lambda_p\rho\big(\lvert[\boldsymbol{W}\boldsymbol{x}_k]^{(p)}\rvert+\varepsilon\big)^{\rho-1}$

Table 1. Examples of the composite function R defined in (2). $\boldsymbol{W}\in\mathbb{R}^{P\times N}$, $[\cdot]^{(p)}$ selects the p-th component of its argument, and $\varepsilon>0$. The last column gives the weights used in Algorithm 1 to define function L_k as per (5).

2. PROPOSED APPROACH

2.1. Minimization problem

We consider problem (1) with

$$(\forall \boldsymbol{x}\in\mathbb{R}^N)\quad R(\boldsymbol{x})=\sum_{p=1}^{P}\lambda_p\phi_p\circ\psi_p(\boldsymbol{x}), \tag{2}$$

where, for every $p\in\{1,\dots,P\}$, $\lambda_p>0$, $\psi_p\colon\mathbb{R}^N\to[0,+\infty]$ is proper, lower semi-continuous and Lipschitz continuous on its domain, and $\phi_p\colon[0,+\infty]\to]-\infty,+\infty]$ is a concave, strictly increasing and twice differentiable function ($\mathcal{C}^2$) on the domain of ψ_p. Particular cases of the function R are provided in Table 1.

2.2. Proposed method

To solve problem (1)-(2), we develop an approximated FB algorithm. Traditional FB algorithm [22, 23] would alternate, at each iteration $k\in\mathbb{N}$, between a gradient step on the Lipschitz-differentiable function F and a proximity step on R. Specifically, given the current iterate $\boldsymbol{x}_k$, the next iterate is defined as

$$\boldsymbol{x}_{k+1}\in\operatorname{prox}_{\gamma_k R}\big(\boldsymbol{x}_k-\gamma_k\nabla F(\boldsymbol{x}_k)\big), \tag{3}$$

where $\gamma_k>0$ and $\operatorname{prox}_{\gamma_k R}$ is the proximity operator of $\gamma_k R$. For every $\widetilde{\boldsymbol{x}}\in\mathbb{R}^N$, the proximity operator of R at $\widetilde{\boldsymbol{x}}$ is defined to be the set of minimizers of $R+\frac{1}{2}\|\cdot-\widetilde{\boldsymbol{x}}\|^2$ [24]. Due to the specific composite form of R, this standard FB approach might not be optimal, and alternative methods are to be investigated.

On the same flavour as reweighting methods, we propose to approximate R by a majorant function. We observe that, for a given iterate $\boldsymbol{x}_k$ and for every $\boldsymbol{x}\in\mathbb{R}^N$,

$$\begin{cases} q_p(\boldsymbol{x},\boldsymbol{x}_k)=\lambda_p\phi_p\circ\psi_p(\boldsymbol{x}_k)+\lambda_{p,k}\big(\psi_p(\boldsymbol{x})-\psi_p(\boldsymbol{x}_k)\big)\\ \lambda_{p,k}=\lambda_p\dot{\phi}_p\circ\psi_p(\boldsymbol{x}_k), \end{cases} \tag{4}$$

where $\dot{\phi}_p$ denotes the first derivative of ϕ_p, is a majorant function of $\lambda_p\phi_p\circ\psi_p$ at $\boldsymbol{x}_k$ (as it is the tangent function of the concave differentiable function $\lambda_p\phi_p$ at $\psi_p(\boldsymbol{x}_k)$). We can deduce from (4) a majorant function of R at $\boldsymbol{x}_k$, given for every $\boldsymbol{x}\in\mathbb{R}^N$ by $Q(\boldsymbol{x},\boldsymbol{x}_k)=\sum_{p=1}^{P}q_p(\boldsymbol{x},\boldsymbol{x}_k)$, and we replace the proximity operator of R in (3) by the proximity operator of $Q(\cdot,\boldsymbol{x}_k)$. Noticing that $\operatorname{prox}_{Q(\cdot,\boldsymbol{x}_k)}=\operatorname{prox}_{L_k}$ where

$$(\forall \boldsymbol{x}\in\mathbb{R}^N)\quad L_k(\boldsymbol{x})=\sum_{p=1}^{P}\lambda_{p,k}\psi_p(\boldsymbol{x}), \tag{5}$$

we propose to solve problem (1)-(2) using Algorithm 1.

Algorithm 1 Reweighted forward-backward algorithm

Initialization: Let $\boldsymbol{x}_0\in\operatorname{dom}R$. Let, for every $k\in\mathbb{N}$, $I_k\in\mathbb{N}^*$, and $(\gamma_{k,i})_{0\leqslant i\leqslant I_k-1}\in]0,+\infty[^{I_k}$.
Iterations:
For $k=0,1,\dots$
  $\widetilde{\boldsymbol{x}}_{k,0}=\boldsymbol{x}_k$,
  for $i=0,\dots,I_k-1$
    $\widetilde{\boldsymbol{x}}_{k,i+1}=\operatorname{prox}_{\gamma_{k,i}L_k}\Big(\widetilde{\boldsymbol{x}}_{k,i}-\gamma_{k,i}\nabla F(\widetilde{\boldsymbol{x}}_{k,i})\Big)$,
  $\boldsymbol{x}_{k+1}=\widetilde{\boldsymbol{x}}_{k,I_k}$.

In Algorithm 1, we can notice that the approximation of R is kept fixed for $I_k\in\mathbb{N}^*$ FB iterations. On the one hand, when $I_k=1$, the approximation L_k is updated for each proximity step. On the other hand, in the limit case when $I_k\to\infty$, the inner-loop is it-self a FB algorithm minimizing $F+L_k$, where k is fixed. According to [22,23], under technical assumptions, these iterations converge to a critical point of $F+L_k$. Thus, in this limit case, Algorithm 1 is very similar to the reweighted methods proposed in [13–16], where each sub-problem must be solved accurately.

2.3. Convergence results

The convergence of proposed approximated forward-backward algorithm is guaranteed by the following result, deduced from [21, Thm. 4.11].

Theorem 2.1 *We consider problem* (1)-(2)*, where F is a Lipschitz-differentiable function with Lipschitz constant*

$\mu > 0$, and R is defined in Section 2.1. Let $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1. Assume that

(i) *H is a coercive function, i.e.* $\lim_{\|\boldsymbol{x}\|\to+\infty} H(\boldsymbol{x}) = +\infty$.

(ii) *The function H is semi-algebraic*[1].

(iii) *There exists $(\underline{\gamma}, \overline{\gamma}) \in]0, +\infty[^2$ such that, for every $k \in \mathbb{N}$, and $i \in \{0, \ldots, I_k - 1\}$, $\underline{\gamma} \leqslant \gamma_{k,i} \leqslant \mu^{-1}(1 - \overline{\gamma})$.*

(iv) *There exists $\overline{I} \in \mathbb{N}^*$ such that, for every $k \in \mathbb{N}$, $0 < I_k \leqslant \overline{I} < +\infty$.*

Then, the sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ converges to critical point $\boldsymbol{x}^\dagger$ of Problem (1)-(2). *Moreover, $(H(\boldsymbol{x}_k))_{k\in\mathbb{N}}$ monotonically converges to $H(\boldsymbol{x}^\dagger)$.*

2.4. Additional remarks

Preconditioned version. As highlighted in [17], and similarly to many first-order minimization methods, the FB algorithm may converge slowly in practice. To overcome this issue, in [21] we have developed a preconditioned version of Algorithm 1, based on a Majorize-Minimize approach. This technique has shown to accelerate drastically the convergence of the iterates [23,25].
Inexact proximity operator. Although the proximity operator computation is simplified by the use of L_k instead of R, some sophisticated choices of functions $(\psi_p)_{1\leqslant p\leqslant P}$ may necessitate the use of sub-iterations. In this case, Theorem 2.1 still holds [21].

3. APPLICATION TO ASTRONOMICAL IMAGING

3.1. Fourier imaging

Computational imaging often involves to solve an inverse problem, where the objective is to find an estimate $\boldsymbol{x}^\dagger \in \mathbb{R}^N$ of an original unknown image $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ from a degraded and/or incomplete observation of this image. For Fourier imaging, the observations $\boldsymbol{y} \in \mathbb{C}^M$ correspond to under-sampled Fourier measurements of the form $\boldsymbol{y} = \boldsymbol{\Phi}\overline{\boldsymbol{x}} + \boldsymbol{b}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ models the under-sampled Fourier operator, and $\boldsymbol{b} \in \mathbb{C}^M$ is a realization of a random variable. More precisely, $\boldsymbol{\Phi}$ is the multiplication of three operators $\boldsymbol{\Phi} = \boldsymbol{GFZ}$, where $\boldsymbol{Z} \in \mathbb{R}^{K\times N}$ is a zero-padding and scaling operator, $\boldsymbol{F} \in \mathbb{C}^{K\times K}$ is the 2D Fourier transform, and $\boldsymbol{G} \in \mathbb{C}^{M\times K}$ is an interpolation matrix. Each row of $\boldsymbol{G}$ contains compact support kernels modelling the non-uniform Fourier transform, centred at the selected frequency.

[1]Semi-algebraicity is a property satisfied by a wide class of functions, which means that their graph is a finite union of sets defined by a finite number of polynomial inequalities. In particular, it is satisfied for the examples provided in Table 1, for standard numerical implementations of the log function.

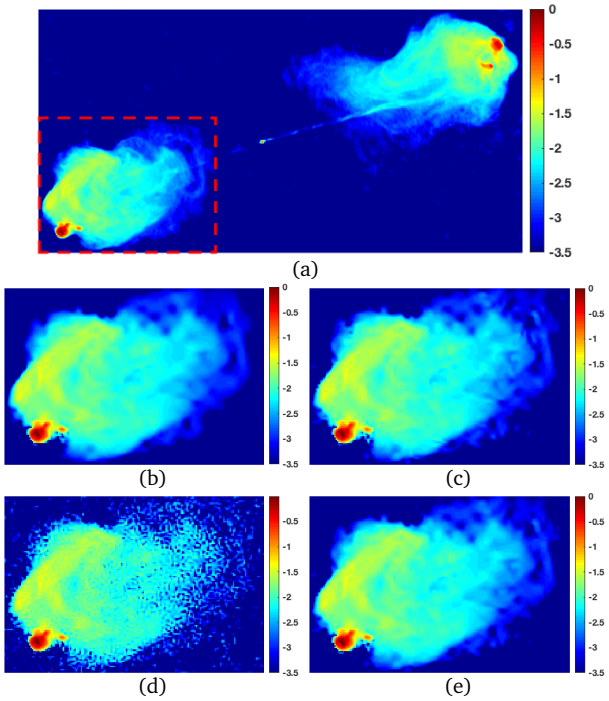

Fig. 1. (a) Original image $\overline{\boldsymbol{x}}$ of Cygnus A of size $N = 256 \times 512$ in log scale. (b), (c) and (d) Zoom on $\boldsymbol{x}^\dagger$ obtained with Algorithm 1 with $I = 1$, $I = 20$ and $I = 90$, respectively. (e) Zoom on $\boldsymbol{x}^\dagger$ obtained with a FB algorithm applied to the log-sum penalty function.

3.2. Application to radio-astronomical imaging

In the remainder of this work we will focus on the application of Fourier imaging in radio astronomy [26]. In this context we assume that $\boldsymbol{b}$ is a realization of an i.i.d. complex-valued random Gaussian variable with standard deviation σ. In addition, during the last decade, reweighted ℓ_1 methods have shown to outperform state-of-the-art approaches in terms of quality for radio-astronomical imaging [27–29]. The reweighted ℓ_1 approach consists in minimising a log-sum penalty term. Therefore, we define the estimate $\boldsymbol{x}^\dagger$ of $\overline{\boldsymbol{x}}$ to be a minimizer of $H = F + R$ with, for every $\boldsymbol{x} \in \mathbb{R}^N$, $F(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{\Phi x} - \boldsymbol{y}\|^2$, and

$$R(\boldsymbol{x}) = \lambda \sum_{c=1}^{C} \log\left(|[\boldsymbol{W x}]^{(c)}| + \varepsilon\right) + \iota_{[0,+\infty[^N}(\boldsymbol{x}), \quad (6)$$

where $\lambda > 0$, $\varepsilon > 0$, and $\boldsymbol{W} \in \mathbb{R}^{C\times N}$ models a sparsity basis (i.e. wavelet transform). Moreover, in (6), $\iota_{[0,+\infty[^N}$ denotes the indicator function of $[0, +\infty[^N$, which is equal to 0 if its argument belongs to $[0, +\infty[^N$, and to $+\infty$ otherwise. As proposed in [28], we choose to promote average sparsity by defining $\boldsymbol{W}$ to be the concatenation of nine sparsity bases: the first eight

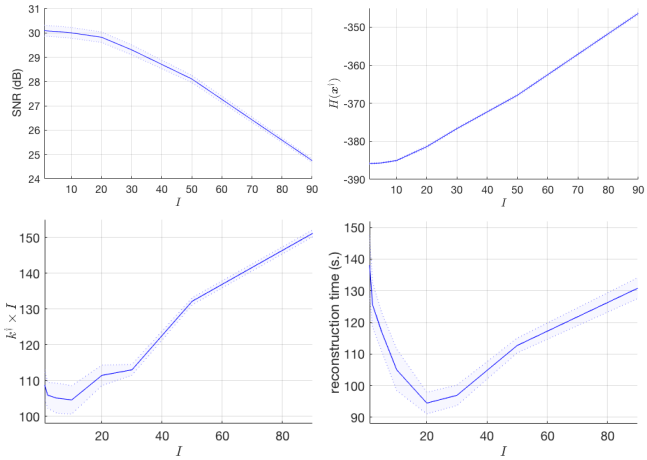

Fig. 2. Simulation results obtained with the proposed method, considering multiple choices for $I_k \equiv I \in \{1, \dots, 90\}$. The continuous lines show the average values and the shadowed areas give the associated standard deviations.

Daubechies wavelet transforms and the Dirac basis. This approach is called "Sparsity Averaging Reweighting Analysis"(SARA).

To minimize H, we propose to use Algorithm 1. To this end, we need to define, for every $k \in \mathbb{N}$, functions L_k as per equation (5), i.e., for every $\boldsymbol{x} \in \mathbb{R}^N$,

$$L_k(\boldsymbol{x}) = \|\Lambda_k \boldsymbol{W} \boldsymbol{x}\|_1 + \iota_{[0,+\infty[^N}(\boldsymbol{x}), \tag{7}$$

where $\Lambda_k = \mathrm{Diag}\left((\lambda_{c,k})_{1 \leqslant c \leqslant C}\right) \in \mathbb{R}^{C \times C}$ is the diagonal weighting matrix whose diagonal coefficients $(\lambda_{c,k})_{1 \leqslant c \leqslant C}$ are given in the first row of Table 1. Although the proximity operator of L_k does not have a closed form solution, it can be computed efficiently using, e.g. a dual FB algorithm [30].

Simulations have been performed on MATLAB. The objective is to show the advantages of the proposed method, mainly of not solving accurately each sub-problem, compared to state-of-the-art methods holding convergence guarantees. Algorithm 1 is considered to have converged when the relative changes between consecutive iterates is smaller than 10^{-4}, or when the relative changes of the objective value is smaller than 10^{-3}.

We consider an image of Cygnus A of size $N = 256 \times 512$, given in Fig. 1(a). The observations consist of $M = N/2$ measurements selected randomly through Gaussian sampling with zero mean and variance 0.25 of the maximum frequency, creating a concentration of data in the centre of the Fourier plane, for low frequencies. The noise level is fixed with iSNR $= 20 \log_{10}\left(\|\boldsymbol{\Phi}\overline{\boldsymbol{x}}\|/(\sigma M)\right) = 20$ dB. The reconstructions are repeated 20 times, changing both the sampling distribution and the noise realization. We run our method fixing the number of iterations for the inner-loop I_k throughout the iterations $k \in \mathbb{N}$, and considering different values for $I_k \equiv I \in \{1, \dots, 90\}$.

Results are reported in Fig. 2. Top-left image shows the SNR $= 20 \log_{10}\left(\|\overline{\boldsymbol{x}}\|/\|\overline{\boldsymbol{x}} - \boldsymbol{x}^\dagger\|\right)$ (in dB) of the image reconstructed by Algorithm 1 for the different values of I (blue curve). The reconstruction quality decreases when I increases. Top-right image shows the value of the objective function H evaluated at the point estimate $\boldsymbol{x}^\dagger$, as a function of I. The best minimum value is obtained when $I = 1$, and $H(\boldsymbol{x}^\dagger)$ increases with I. The second row shows computation performance of the method: the bottom-left image shows the total number of iterations needed to reach convergence versus I, and the bottom-right image shows the reconstruction time in sec. as a function of I. We can notice that the optimal value it terms of computation time is $I = 20$. This value appears to be the best compromise in this example as it does not affect the reconstruction quality. For visual comparison a zoom on the reconstructed images obtained with the proposed method for $I = 1$, $I = 20$ and $I = 90$, are provided in Fig. 1(b), (c), and (d), respectively.

For the sake of completeness, we also implemented a basic FB algorithm [22], where the proximity operator of R is computed with a dual FB algorithm, using the implementation of the proximity operator of the log-sum function provided in `http://proximity-operator.net` (note that the convergence guarantees of the dual FB do not hold due to the non-convexity of the log-sum function). The obtained results have similar quality as the proposed approach when $I \leqslant 20$ (SNR $= 30$ dB and $H(\boldsymbol{x}^\dagger) = -380$), however it requires a much higher number of iterations to converge (average of 2000 iterations) with an average of 30 minutes. However for $I > 20$, this method provides better reconstruction results than Algorithm 1 (with a much higher computational cost), and thus to state-of-the-art reweighted methods when the sub-problems must be solved accurately. Fig. 1(e) shows a zoom on the images obtained with this method.

4. CONCLUSION

In this work we have presented a FB algorithm designed to minimize composite functions, through a novel reweighted procedure. The proposed reweighted FB algorithm incorporates the computation of the weights directly in the iterations. Unlike state-of-the-art reweighted methods, it does not necessitate to solve accurately multiple minimization sub-problems. Our algorithm benefits from convergence guarantees very similar to the basic FB algorithm in a non-convex context. Simulations on a radio-astronomical imaging problem have been performed. Results highlight the advantage of re-computing the weights without waiting for the convergence of the sub-problems (unlike state-of-the-art approaches) both in terms of reconstruction quality and computation time.

5. REFERENCES

[1] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2007.

[2] A. Auria, A. Daducci, J.P. Thiran, and Y. Wiaux, "Structured sparsity through reweighting and application to diffusion mri," in *IEEE European Signal Processing Conference (EUSIPCO 2015)*. IEEE, 2015, pp. 454–458.

[3] O. Taheri and S. A. Vorobyov, "Reweighted l1-norm penalized lms for sparse channel estimation and its analysis," *Signal Processing*, vol. 104, pp. 70–79, 2014.

[4] L. Weizman, Y. C. Eldar, and D. Ben-Bashat, "Compressed sensing for longitudinal mri: An adaptive-weighted approach," *Med. Phys.*, vol. 42, no. 9, pp. 5195–5207, 2015.

[5] H-M Jeon, J.-Y. Lee, G.-M. Jeong, and S.-I. Choi, "Data reconstruction using iteratively reweighted l1-principal component analysis for an electronic nose system," *PLOS ONE*, vol. 13, no. 7, pp. 1–19, 07 2018.

[6] E. J. Candès et al., "Compressive sampling," in *Proceedings of the international congress of mathematicians*. Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[7] J. V. Burke, "Descent methods for composite nondifferentiable optimization problems," *Math. Program.*, vol. 33, no. 3, pp. 260–279, 1985.

[8] C. Cartis, N. I. M. Gould, and P. L. Toint, "On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1721–1739, 2011.

[9] R. Fletcher, "A model algorithm for composite nondifferentiable optimization problems," in *Nondifferential and Variational Techniques in Optimization*, pp. 67–76. Springer, 2009.

[10] A. S. Lewis and S. J. Wright, "A proximal method for composite minimization," *Math. Program.*, vol. 158, pp. 501–546, 2015.

[11] M. J. D. Powell, "General algorithms for discrete nonlinear approximation calculations," in *Approximation theory, IV*, pp. 187–218. Academic Press, New York, 1983.

[12] S. J. Wright, "Convergence of an inexact algorithm for composite nonsmooth optimization," *IMA J. Numer. Anal.*, vol. 10, no. 3, pp. 299–321, 1990.

[13] D. Drusvyatskiy, A.d. Ioffe, and A. S. Lewis, "Nonsmooth optimization using taylor-like models: error bounds, convergence, and termination criteria," Tech. Rep., 2016, arXiv:1610.03446.

[14] J. Geiping and M. Moeller, "Composite optimization by nonconvex majorization-minimization," *SIAM J. Imaging Sci.*, vol. 11, no. 4, pp. 2494–2598, 2018.

[15] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM J. Imaging Sci.*, vol. 8, no. 1, pp. 331–372, 2015.

[16] P. Ochs, J. Fadili, and T. Brox, "Non-smooth nonconvex bregman minimization: Unification and new algorithms," *J. Optim. Theory. Appl.*, vol. 181, no. 1, pp. 244–278, 2019.

[17] G. H.-G. Chen and R. T. Rockafellar, "Convergence rates in forward-backward splitting," *SIAM J. Optim.*, vol. 7, no. 2, pp. 421–444, 1997.

[18] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[19] P.-L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1079.

[20] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control Optim.*, vol. 38, pp. 431–446, 2000.

[21] A. Repetti and Y. Wiaux, "Variable metric forward-backward algorithm for composite minimization problems," Tech. Rep., Jul. 2019, arXiv:1907.11486.

[22] 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. Program.*, vol. 137, pp. 91–129, Feb. 2011.

[23] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function," *J. Optim. Theory Appl.*, vol. 162, no. 1, Jul. 2014.

[24] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.

[25] A. Repetti, E. Chouzenoux, and J.-C. Pesquet, "A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems," in *Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, 4-9 May 2014, pp. 1498–1502.

[26] A. R. Thompson, "Interferometry and synthesis in radio astronomy," 2001.

[27] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst, "Compressed sensing for radio interferometry: prior-enhanced basis pursuit imaging techniques," in *SPARS" 09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[28] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-P. Thiran, J.-C. Pesquet, and Y. Wiaux, "Scalable splitting algorithms for big-data interferometric imaging in the ska era," *Mon. Not. R. Astron. Soc.*, vol. 462, no. 4, pp. 4314–4335, 2016.

[29] A. Dabbech, A. Onose, A. Abdulaziz, A. Perley, O. M. Smirnov, and Y. Wiaux, "Cygnus a super-resolved via convex optimization from vla data," *Mon. Not. R. Astron. Soc.*, vol. 476, no. 3, pp. 2853–2866, 2018.

[30] P. L. Combettes, D. Dũng, and B. C. Vũ, "Proximity for sums of composite functions," *J. Math. Anal. Appl.*, vol. 380, no. 2, pp. 680–688, Aug. 2011.