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 $\mathbf{x}^\dagger$ of the unknown object of interest $\mathbf{x} \in \mathbb{R}^N$, 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 $\mathbf{x}$:

$$\mathbf{x}^\dagger \in \text{Argmin}_{\mathbf{x} \in \mathbb{R}^N} \{H(\mathbf{x}) = F(\mathbf{x}) + R(\mathbf{x})\}. \quad (1)$$

We focus on the case where $F: \mathbb{R}^N \to \mathbb{R}$ is Lipschitz-differentiable, and $R: \mathbb{R}^N \to [\infty, +\infty]$ 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 [1,2,3,4,5]. They consist in promoting sparsity by mimicking the $\ell_0$ 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 $\ell_1$ minimization approach [6], aiming to minimize a log-sum prior by successively solving weighted-$\ell_1$ sub-problems. The convergence of reweighted methods have been investigated during the last decades, e.g., in [7,8,9,10,11,12,13,14,15,16]. 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 [17,18,19,20], 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 reweighted FB algorithm. Simulation results are given in Section 3. Finally, we conclude...
Table 1. Examples of the composite function $R$ defined in (2). \( W \in \mathbb{R}^{P \times N} \), \( \cdot^{(p)} \) 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).

<table>
<thead>
<tr>
<th>Penetration</th>
<th>$\psi_p(x)$</th>
<th>$\phi_p(u)$</th>
<th>$R(x)$</th>
<th>$\lambda_{p,k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-sum</td>
<td>$</td>
<td>Wx</td>
<td>^{(p)}$</td>
<td>$(u + \varepsilon)$</td>
</tr>
<tr>
<td>Cauchy</td>
<td>$</td>
<td>Wx</td>
<td>^{(p)}^2$</td>
<td>$(u + \varepsilon)$</td>
</tr>
<tr>
<td>Semi-smooth $\ell_p$-norm, $0 &lt; p &lt; 1$</td>
<td>$</td>
<td>Wx</td>
<td>^{(p)}$</td>
<td>$(u + \varepsilon) - \varepsilon^p$</td>
</tr>
</tbody>
</table>

2. PROPOSED APPROACH

2.1. Minimization problem

We consider problem (1) with

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

where, for every $p \in \{1, \ldots, P\}$, $\lambda_p > 0$, $\psi_p : \mathbb{R}^N \to [0, +\infty)$ is proper, lower semi-continuous and Lipschitz continuous on its domain, and $\phi_p : [0, +\infty) \to (-\infty, +\infty]$ is a concave, strictly increasing and differentiable function, such that $\phi_p \circ \psi_p$ is Lipschitz-continuous on the domain of $\psi_p$, where $\phi_p$ denotes the first derivative of $\phi_p$. Particular cases of the function $R$ are provided in Table 1.

2.2. Proposed method

To solve problem (1)-(2), we develop an approximated forward-backward 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 $x_k$, the next iterate is defined as

\[
x_{k+1} \in \text{prox}_{\gamma_k R}(x_k - \gamma_k \nabla F(x_k)), \tag{3}
\]

where $\gamma_k > 0$ and $\text{prox}_{\gamma_k R}$ is the proximity operator of $\gamma_k R$. For every $x \in \mathbb{R}^N$, the proximity operator of $R$ at $x$ is defined as the set of minimizers of $R + \frac{1}{\gamma_k} \| : -x \|^2$ \[24\]. 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 $x_k$ and for every $x \in \mathbb{R}^N$,

\[
\begin{align*}
q_p(x, x_k) &= \lambda_p \phi_p \circ \psi_p(x_k) + \lambda_{p,k} (\psi_p(x) - \psi_p(x_k)) \\
\lambda_{p,k} &= \lambda_p \phi_p \circ \psi_p(x_k),
\end{align*}
\]

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

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

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

In Algorithm 1, we can notice that the approximation of $R$ is kept fixed for $k \in \mathbb{N}$ during FB iterations. On the one hand, when $k = 1$, the approximation $L_k$ is updated for each proximity step. On the other hand, in the limit case when $k \to \infty$, the inner-loop is itself 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 \[13, 14, 15, 16\], 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.

---

Algorithm 1 Reweighted forward-backward algorithm

**Initialization:** Let $x_0 \in \text{dom } R$. Let, for every $k \in \mathbb{N}$, $I_k \in \mathbb{N}^*$, and $(\gamma_{k,i})_{0 \leq i \leq I_k - 1} \in [0, +\infty]^{I_k}$.

**Iterations:**

For $k = 0, 1, \ldots$,

\[
\begin{align*}
\bar{x}_{k,0} &= x_k, \\
\text{for } i = 0, \ldots, I_k - 1, \\
\bar{x}_{k,i+1} &= \text{prox}_{\gamma_{k,i} L_k} \left( \bar{x}_{k,i} - \gamma_{k,i} \nabla F(\bar{x}_{k,i}) \right), \\
x_{k+1} &= \bar{x}_{k,I_k}.
\end{align*}
\]
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 $(x_k)_{k \in \mathbb{N}}$ be a sequence generated by Algorithm 1. Assume that

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

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

(iii) There exists $T \in \mathbb{N}^+$ such that, for every $k \in \mathbb{N}$, $0 < I_k \leq T < +\infty$.

(iv) The functions $F$, $\psi_1$, $\ldots$, $\psi_P$ are semi-algebraic. A function is semi-algebraic if its graph is a finite union of sets defined by a finite number of polynomial inequalities.

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

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 $(\psi_p)_{1 \leq p \leq P}$ 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 $x^* \in \mathbb{R}^N$ of an original unknown image $\mathbf{x} \in \mathbb{R}^N$ from a degraded and/or incomplete observation of this image. For Fourier imaging, the observations $y \in \mathbb{C}^M$ correspond to under-sampled Fourier measurements of the form $y = \Phi \mathbf{x} + b$, where $\Phi \in \mathbb{C}^{M \times N}$ models the under-sampled Fourier operator, and $b \in \mathbb{C}^M$ is a realization of a random variable. More precisely, $\Phi$ is the multiplication of three operators $\Phi = GFZ$, where $Z \in \mathbb{R}^{K \times N}$ is a zero-padding and scaling operator, $F \in \mathbb{C}^{K \times K}$ is the 2D Fourier transform, and $G \in \mathbb{C}^{M \times K}$ is an interpolation matrix. Each row of $G$ contains compact support kernels modelling the non-uniform Fourier transform, centred at the selected frequency.

![Fig. 1.](image-url)
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 $x \in \mathbb{R}^N$,

$$L_k(x) = \|\Lambda_k Wx\|_1 + I_{[0, +\infty]}(x), \quad (7)$$

where $\Lambda_k = \text{Diag}((\lambda_{c,k})_{1 \leq c \leq C}) \in \mathbb{R}^{C \times C}$ is the diagonal weighting matrix whose diagonal coefficients $(\lambda_{c,k})_{1 \leq c \leq C}$ 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 $\text{SNR} = 20 \log_{10} (\|\Phi x\|/(\sigma M)) = 20$ 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 $I_k \equiv I \in \{1, \ldots, 90\}$.

Results are reported in Fig. 2. Top-left image shows the SNR $= 20 \log_{10} (\|x^\dagger/\|x^\dagger - x^\dagger\|)$ (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 $x^\dagger$, as a function of $I$. The best minimum value is obtained when $I = 1$, and $H(x^\dagger)$ 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](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 $I \leq 20$ (SNR $= 30$ dB and $H(x^\dagger) = -380$), 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


