Journal of Physics: Conference Series **756** (2016) 012001 doi:10.1088/1742-6596/756/1/012001

On the constrained minimization of smooth Kurdyka–Łojasiewicz functions with the scaled gradient projection method

Marco Prato[1], **Silvia Bonettini**[2], **Ignace Loris**[3], **Federica Porta**[2] **and Simone Rebegoldi**[1]

[1] Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/b, 41125 Modena, Italy

[2] Dipartimento di Matematica e Informatica, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy

[3] Département de Mathématique, Université Libre de Bruxelles, Boulevard du Triomphe, 1050 Bruxelles, Belgium

E-mail: marco.prato@unimore.it

Abstract. The scaled gradient projection (SGP) method is a first-order optimization method applicable to the constrained minimization of smooth functions and exploiting a scaling matrix multiplying the gradient and a variable steplength parameter to improve the convergence of the scheme. For a general nonconvex function, the limit points of the sequence generated by SGP have been proved to be stationary, while in the convex case and with some restrictions on the choice of the scaling matrix the sequence itself converges to a constrained minimum point. In this paper we extend these convergence results by showing that the SGP sequence converges to a limit point provided that the objective function satisfies the Kurdyka–Łojasiewicz property at each point of its domain and its gradient is Lipschitz continuous.

1. The scaled gradient projection method

The scaled gradient projection algorithm (SGP) [4] belongs to the class of first-order methods designed to solve any constrained optimization problem of the form

$$\min_{x\in\mathbb{R}^n} \Psi(x) \equiv f(x) + \iota_\Omega(x), \tag{1}$$

where f is a continuously differentiable function and ι_Ω is the indicator function of the nonempty, closed and convex set Ω. In particular, the $(k+1)$–th SGP iteration is computed as

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} = x^{(k)} + \lambda_k(\underbrace{P_{\Omega,D_k^{-1}}(x^{(k)} - \alpha_k D_k \nabla f(x^{(k)}))}_{=:y^{(k)}} - x^{(k)}),$$

where α_k is a scalar steplength parameter, D_k is a symmetric positive definite matrix and $P_{\Omega,D_k^{-1}}(\cdot)$ is the projection onto Ω associated to the norm induced by D_k^{-1} [4]. The step along the descent direction $d^{(k)}$ is performed by means of the linesearch parameter $\lambda_k = \delta^{m_k}$, where $\delta \in (0, 1)$ and m_k is the smallest non-negative integer such that the monotone Armijo condition

$$f(x^{(k)} + \lambda_k d^{(k)}) \leq f(x^{(k)}) + \beta\lambda_k \nabla f(x^{(k)})^T d^{(k)} \tag{2}$$

is satisfied for a fixed value of the parameter $\beta \in (0, 1)$.
The SGP method is a variable metric forward-backward algorithm [11, 12] which has been exploited in the last years for the solution of different real-world inverse problems [5, 6, 7, 16, 17, 18]. The main difference between SGP and the standard forward-backward schemes is the presence of two independent parameters α_k and λ_k with a complete different role: while the last one is automatically computed with the Armijo condition (2) to guarantee the sufficient decrease of the objective function, the first one can be chosen to improve the actual convergence rate of the method, exploiting thirty years of literature in numerical optimization [3, 13, 19, 20]. We also remark that, unlike the schemes presented e.g. in [2, Sect. 5], in SGP the explicit knowledge of the gradient Lipschitz constant is not required.
The general convergence result on the SGP sequence states that all its limit points are stationary for problem (1), provided that both the steplength α_k and the eigenvalues of D_k are chosen in prefixed positive intervals $[\alpha_{\min}, \alpha_{\max}]$ and $[\frac{1}{\mu}, \mu]$, respectively [4]. Convergence of the sequence to a minimum point of (1) has recently been proved for convex objective functions by choosing suitable adaptive bounds for the eigenvalues of the scaling matrices [8]. In the following, we will consider a modified version of SGP in which, at each iteration $k \in \mathbb{N}$, we compute

$$x^{(k+1)} = \begin{cases} y^{(k)} & \text{if} \quad f(y^{(k)}) < f(x^{(k)} + \lambda_k d^{(k)}) \\ x^{(k)} + \lambda_k d^{(k)} & \text{otherwise.} \end{cases} \tag{3}$$

so that the sequence $\{f(x^{(k)})\}_{k\in\mathbb{N}}$ is forced to assume lower values than it originally would with SGP. By doing so, the new sequence satisfies the condition $f(x^{(k+1)}) \leq f(x^{(k)} + \lambda_k d^{(k)})$, which still guarantees the stationarity of the limit points [8]. The aim of this paper is to prove the convergence of this modified SGP scheme if the (nonconvex) objective function Ψ satisfies the Kurdyka–Łojasiewicz (KL) property [15, 14], which holds true for most of the functions commonly used in inverse problems as p norms, Kullback-Leibler divergence and indicator functions of box plus equality constraints. A generalization of the proposed method and the related convergence proof to the minimization of the sum of smooth (nonconvex) and convex (nonsmooth) KL functions is in progress [10].

2. Preliminary results

Let $\Psi : \mathbb{R}^n \longrightarrow \mathbb{R}\cup\{+\infty\}$ be a proper, lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, let us set $[\eta_1 < \Psi < \eta_2] = \{z \in \mathbb{R}^n : \eta_1 < \Psi(z) < \eta_2\}$. Moreover, we denote with $\partial\Psi(z)$ the subdifferential of Ψ at $z \in \mathbb{R}^n$ and with $\text{dist}(z, \Omega)$ the distance between a point z and a set $\Omega \subset \mathbb{R}^n$. The function Ψ is said to have the KL property at $\overline{z} \in \text{dom}\partial\Psi := \{z \in \mathbb{R}^n : \partial\Psi(z) \neq \emptyset\}$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of $\overline{z}$ and a continuous concave function $\varphi : [0, \eta) \longrightarrow [0, +\infty)$ such that $\varphi(0) = 0$, φ is C^1 on $(0, \eta)$, $\varphi'(s) > 0$ for all $s \in (0, \eta)$ and the KL inequality

$$\varphi'(\Psi(z) - \Psi(\overline{z}))\text{dist}(0, \partial\Psi(z)) \geq 1. \tag{4}$$

holds for all $z \in U \cap [\Psi(\overline{z}) < \Psi < \Psi(\overline{z}) + \eta]$. If Ψ satisfies the KL property at each point of $\text{dom}\partial\Psi$, then Ψ is called a KL function.
The convergence proof of SGP for KL functions follows the ideas presented in [2], in which the authors proved an abstract convergence result for descent methods satisfying the following three conditions:

H1. (*Sufficient decrease condition*). There exists $a > 0$ such that $\Psi(x^{(k+1)}) + a\|x^{(k+1)} - x^{(k)}\|^2 \leq \Psi(x^{(k)})$ for each $k \in \mathbb{N}$.

H2. (*Relative error condition*). There exists $b > 0$ and, for each $k \in \mathbb{N}$, $w^{(k+1)} \in \partial\Psi(x^{(k+1)})$ such that $\|w^{(k+1)}\| \leq b\|x^{(k+1)} - x^{(k)}\|$.

H3. (*Continuity condition*). There exists a subsequence $\{x^{(k_j)}\}_{j\in\mathbb{N}}$ and $\tilde{x}$ such that $x^{(k_j)} \to \tilde{x}$ and $\Psi(x^{(k_j)}) \to \Psi(\tilde{x})$, as $j \to \infty$.

In our case, condition H3 is assured by the continuity of Ψ in Ω and the fact that $x^{(k)} \in \Omega$, for every $k \in \mathbb{N}$. Indeed H3 is needed in [2] only to ensure the stationarity of the limit point $\tilde{x}$, which has already been proved for SGP in [4].

Throughout the entire section, $\{x^{(k)}\}_{k\in\mathbb{N}}$ will denote the sequence generated by SGP. The following lemma (whose proof follows from [9, Lemma 4.1, Lemma 4.3 and Proposition 4.2]) and corollary guarantee that condition H1 holds for SGP.

Lemma 1 *The sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ satisfies*

$$\frac{1}{2\mu\alpha_{\max}}\|y^{(k)} - x^{(k)}\|^2 \leq -\nabla f(x^{(k)})^T(y^{(k)} - x^{(k)}). \tag{5}$$

Moreover, if ∇f is L-Lipschitz continuous on Ω, then:

- *there exists $c > 0$ such that for all $\lambda \in [0, 1]$*
$$f(x^{(k)} + \lambda d^{(k)}) \leq f(x^{(k)}) + \lambda(1 - cL\lambda)\nabla f(x^{(k)})^T d^{(k)}; \tag{6}$$
- *the sequence $\{\lambda_k\}_{k\in\mathbb{N}}$ of SGP linesearch parameters is bounded away from zero, i.e.,*
$$\lambda_k \geq \lambda_{\min}\ \forall k \in \mathbb{N}, \quad \lambda_{\min} > 0. \tag{7}$$
- *if f is also bounded from below, then $0 < -\sum_{k=0}^{+\infty} \nabla f(x^{(k)})^T d^{(k)} < +\infty$ and, consequently,*
$$\lim_{k\to+\infty} \nabla f(x^{(k)})^T d^{(k)} = 0. \tag{8}$$

Corollary 1 *If ∇f is L-Lipschitz continuous on Ω, then*

$$\Psi(x^{(k+1)}) + a\|x^{(k+1)} - x^{(k)}\|^2 \leq \Psi(x^{(k)}). \tag{H1}$$

Proof: Combining (5) with the backtracking rule (2) immediately yields

$$f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{\beta\lambda_k}{2\alpha_{\max}\mu}\|y^{(k)} - x^{(k)}\|^2. \tag{9}$$

Because of (3), it is either $x^{(k+1)} = y^{(k)}$ or $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$. In the first case, using (9) with (7) leads to

$$f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{\beta\lambda_{\min}}{2\alpha_{\max}\mu}\|x^{(k+1)} - x^{(k)}\|^2. \tag{10}$$

In the second case, we obtain the same inequality by using $y^{(k)} - x^{(k)} = \frac{1}{\lambda_k}(x^{(k+1)} - x^{(k)})$ and $\lambda_{\min} \leq \lambda_k \leq 1$ in (9). Finally, (H1) follows by adding the indicator function ι_Ω to both terms of (10) and by taking $a = \frac{\beta\lambda_{\min}}{2\alpha_{\max}\mu}$. □

We are now able to prove a slight variant of condition H2 for SGP, in which a subgradient of Ψ in $y^{(k)}$ (and not in $x^{(k)}$) is provided.

Lemma 2 *Suppose that ∇f is L-Lipschitz continuous on Ω. There exist $b > 0$ and $v^{(k)} \in \partial\Psi(y^{(k)})$ for all $k \in \mathbb{N}$ such that*

$$\|v^{(k)}\| \leq b\|x^{(k+1)} - x^{(k)}\|. \tag{H2}$$

Journal of Physics: Conference Series **756** (2016) 012001 doi:10.1088/1742-6596/756/1/012001

Proof: By setting $\tilde{h}_k(y) = \nabla f(x^{(k)})^T(y - x^{(k)}) + \frac{1}{2\alpha_k}\|y - x^{(k)}\|^2_{D_k^{-1}} + \iota_\Omega(y)$, we rewrite $y^{(k)}$ as

$$y^{(k)} = \underset{y\in\mathbb{R}^n}{\arg\min}\ \tilde{h}_k(y) \Leftrightarrow 0 \in \partial\tilde{h}_k(y^{(k)}) \Leftrightarrow 0 \in \nabla f(x^{(k)}) + \frac{1}{\alpha_k}D_k^{-1}(y^{(k)} - x^{(k)}) + \partial\iota_\Omega(y^{(k)})$$
$$\Leftrightarrow w^{(k)} = -\nabla f(x^{(k)}) + \frac{1}{\alpha_k}D_k^{-1}(x^{(k)} - y^{(k)}) \in \partial\iota_\Omega(y^{(k)}).$$

Let us define $v^{(k)} = \nabla f(y^{(k)}) + w^{(k)} \in \partial\Psi(y^{(k)})$ for all $k \in \mathbb{N}$. By using the Lipschitz continuity of ∇f and (7), the following inequalities hold true:

$$\|v^{(k)}\| \le \|\nabla f(y^{(k)}) - \nabla f(x^{(k)})\| + \frac{\mu}{\alpha_k}\|y^{(k)} - x^{(k)}\| \le (L + \frac{\mu}{\alpha_{\min}})\|y^{(k)} - x^{(k)}\|$$
$$\le \frac{1}{\lambda_k}(L + \frac{\mu}{\alpha_{\min}})\|x^{(k+1)} - x^{(k)}\| \le \frac{1}{\lambda_{\min}}(L + \frac{\mu}{\alpha_{\min}})\|x^{(k+1)} - x^{(k)}\|.$$

The thesis holds by choosing $b = \frac{1}{\lambda_{\min}}(L + \frac{\mu}{\alpha_{\min}})$. □

3. Convergence result

In this section we present the convergence proof of SGP, which differs from that in [2, Lemma 2.6] by the presence of the projection step at each $k \in \mathbb{N}$, which was not considered in [2].

Theorem 1 *Suppose Ψ is a KL function and ∇f is L-Lipschitz continuous on Ω. If $\bar{x}$ is a limit point of $\{x^{(k)}\}_{k\in\mathbb{N}}$, then $\bar{x}$ is a stationary point for* (1) *and $x^{(k)}$ converges to $\bar{x}$.*

Proof: The stationarity of $\bar{x}$ has been proved in [4]. Since Ψ is a KL function, it satisfies the KL property at each point of Ω and, in particular, at $\bar{x}$. This means that there exist η, φ and U as in Section 2, such that the KL inequality (4) holds at $\bar{x}$. Equation (H1) implies that $\{\Psi(x^{(k)})\}_{k\in\mathbb{N}}$ is a non increasing sequence and hence $\Psi(x^{(k)}) \to \gamma$, $\Psi(x^{(k)}) \ge \gamma$. By the continuity of Ψ on Ω and the fact that $\bar{x}$ is a limit point of $\{x^{(k)}\}_{k\in\mathbb{N}}$, we deduce that $\gamma = \Psi(\bar{x})$. From (6) with $\lambda = 1$ and by definition (3) of $x^{(k+1)}$ we obtain

$$\Psi(x^{(k+1)}) \le \Psi(y^{(k)}) \le \Psi(x^{(k)}) + (1 - cL)\nabla f(x^{(k)})^T d^{(k)}. \tag{11}$$

Since $\Psi(x^{(k)}) \to \Psi(\bar{x})$ and (8) holds true, from (11) we also have that $\Psi(y^{(k)}) \to \Psi(\bar{x})$. Consequently, for all sufficiently large k we have

$$\Psi(\bar{x}) \le \Psi(x^{(k)}) \le \Psi(y^{(k-1)}) < \Psi(\bar{x}) + \eta. \tag{12}$$

Furthermore, let $\rho > 0$ be such that $B(\bar{x}, \rho) \subset U$. Then using the continuity of φ, one can choose $k_0 \in \mathbb{N}$ such that both (12) and the following technical condition are satisfied:

$$\|\bar{x} - x^{(k_0)}\| + 3\sqrt{\frac{\Psi(x^{(k_0)}) - \Psi(\bar{x})}{a\lambda_{\min}^2}} + \frac{b}{a}\varphi(\Psi(x^{(k_0)}) - \Psi(\bar{x})) < \rho. \tag{13}$$

We will now use $\{x^{(k)}\}_{k\in\mathbb{N}}$ to denote the sequence $\{x^{(k+k_0)}\}_{k\in\mathbb{N}}$. Let us rewrite (H1) as

$$\|x^{(k+1)} - x^{(k)}\| \le \sqrt{\frac{\Psi(x^{(k)}) - \Psi(x^{(k+1)})}{a}}, \tag{14}$$

which, by recalling $x^{(k+1)} - x^{(k)} = \lambda_k(y^{(k)} - x^{(k)})$ and (7), yields also

$$\|y^{(k)} - x^{(k)}\| \le \sqrt{\frac{\Psi(x^{(k)}) - \Psi(x^{(k+1)})}{a\lambda_{\min}^2}}. \tag{15}$$

Journal of Physics: Conference Series **756** (2016) 012001 doi:10.1088/1742-6596/756/1/012001

Fix $k \geq 1$. We state that if $x^{(k)}, y^{(k-1)} \in B(\bar{x}, \rho)$, then

$$2\|x^{(k+1)} - x^{(k)}\| \leq \|x^{(k)} - x^{(k-1)}\| + \varphi_k, \tag{16}$$

where $\varphi_k = \frac{b}{a}[\varphi(\Psi(x^{(k)}) - \Psi(\bar{x})) - \varphi(\Psi(x^{(k+1)}) - \Psi(\bar{x}))]$. Observe that, because of (12), the quantity $\varphi(\Psi(x^{(k)}) - \Psi(\bar{x}))$ makes sense for all $k \in \mathbb{N}$, and thus φ_k is well posed.
If $x^{(k+1)} = x^{(k)}$ inequality (16) holds trivially. Then we assume $x^{(k+1)} \neq x^{(k)}$. This assumption, combined with (14) and (12), guarantees that $x^{(k)}, y^{(k-1)} \in B(\bar{x}, \rho) \cap [\Psi(\bar{x}) < \Psi < \Psi(\bar{x}) + \eta]$ and therefore we can use the KL inequality in $x^{(k)}$ and $y^{(k-1)}$.
By exploiting the KL inequality at $y^{(k-1)}$ with (H2), it follows that $v^{(k-1)} \neq 0$ and $x^{(k-1)} \neq x^k$. Since $v^{(k-1)} \in \partial\Psi(y^{(k-1)})$, we can use again the KL inequality combined with (H2) to obtain

$$\varphi'(\Psi(y^{(k-1)}) - \Psi(\bar{x})) \geq \frac{1}{\|v^{(k-1)}\|} \geq \frac{1}{b\|x^{(k)} - x^{(k-1)}\|}. \tag{17}$$

Since φ is concave, its derivative is non increasing, thus $\Psi(y^{(k-1)}) - \Psi(\bar{x}) \geq \Psi(x^{(k)}) - \Psi(\bar{x})$ implies $\varphi'(\Psi(x^{(k)}) - \Psi(\bar{x})) \geq \varphi'(\Psi(y^{(k-1)}) - \Psi(\bar{x}))$. This fact applied to inequality (17) leads to

$$\varphi'(\Psi(x^{(k)}) - \Psi(\bar{x})) \geq \frac{1}{b\|x^{(k)} - x^{(k-1)}\|}. \tag{18}$$

Then following the same procedure of [2, Lemma 2.6], which uses the concavity of φ, (H1) and (18), we obtain $\|x^{(k+1)} - x^{(k)}\|^2 \leq \varphi_k\|x^{(k)} - x^{(k-1)}\|$ which, by applying the inequality $2\sqrt{uv} \leq u + v$, gives relation (16).
We are now going to show that for $j = 1, 2, \ldots$

$$x^{(j)}, y^{(j-1)} \in B(\bar{x}, \rho), \tag{19}$$

$$\sum_{i=1}^{j} \|x^{(i+1)} - x^{(i)}\| + \|x^{(j+1)} - x^{(j)}\| \leq \|x^{(1)} - x^{(0)}\| + \chi_j, \tag{20}$$

where $\chi_j = \frac{b}{a}[\varphi(\Psi(x^{(1)}) - \Psi(\bar{x})) - \varphi(\Psi(x^{(j+1)}) - \Psi(\bar{x}))]$.
In order to prove (19)–(20), let us reason by induction on j. Using the triangle inequality, (14) with $k = 0$, the monotonicity of $\{\Psi(x^{(k)})\}_{k\in\mathbb{N}}$ and (13) we have

$$\|\bar{x} - x^{(1)}\| \leq \|\bar{x} - x^{(0)}\| + \|x^{(0)} - x^{(1)}\| \leq \|\bar{x} - x^{(0)}\| + \sqrt{\frac{\Psi(x^{(0)}) - \Psi(\bar{x})}{a}} < \rho,$$

namely $x^{(1)} \in B(\bar{x}, \rho)$. Using (15) with $k = 0$ and applying the same arguments as before, we also have $y^{(0)} \in B(\bar{x}, \rho)$. Finally, direct use of (16) shows that (20) holds with $j = 1$.
By induction, suppose that (19)–(20) hold for some $j \geq 1$. Since proving that $x^{(j+1)} \in B(\bar{x}, \rho)$ is identical to [2], we focus on $y^{(j)} \in B(\bar{x}, \rho)$. We rewrite (15) by noticing that $\Psi(\bar{x}) \leq \Psi(x^{(k+1)}) \leq \Psi(x^{(k)}) \leq \Psi(x^{(0)})$, which implies that $\|y^{(j)} - x^{(j)}\| \leq \sqrt{\left(\Psi(x^{(0)}) - \Psi(\bar{x})\right)/a\lambda_{\min}^2}$. By using this last equation, the triangle inequality, (20) with $k = j$ and (13), we have

$$\begin{aligned}\|\bar{x} - y^{(j)}\| &\leq \|\bar{x} - x^{(0)}\| + \|x^{(0)} - x^{(1)}\| + \sum_{i=1}^{j} \|x^{(i+1)} - x^{(i)}\| + \|x^{(j+1)} - x^{(j)}\| + \|x^{(j)} - y^{(j)}\| \\ &\leq \|\bar{x} - x^{(0)}\| + 2\|x^{(0)} - x^{(1)}\| + \chi_j + \|x^{(j)} - y^{(j)}\| \\ &\leq \|\bar{x} - x^{(0)}\| + 3\sqrt{\frac{\Psi(x^{(0)}) - \Psi(\bar{x})}{a\lambda_{\min}^2}} + \frac{b}{a}\varphi(\Psi(x^{(0)}) - \Psi(\bar{x})) < \rho,\end{aligned}$$

or equivalently $y^{(j)} \in B(\bar{x}, \rho)$. Hence, (16) holds with $k = j + 1$, i.e. $2\|x^{(j+2)} - x^{(j+1)}\| \leq \|x^{(j+1)} - x^{(j)}\| + \varphi_{j+1}$. Adding the above inequality with (20) (with $k = j$) yields (20) with $k = j + 1$, which completes the induction proof.

From (20) it immediately follows $\sum_{i=1}^{j} \|x^{(i+1)} - x^{(i)}\| \leq \|x^{(1)} - x^{(0)}\| + \frac{b}{a}\varphi(\Psi(x^{(1)}) - \Psi(\bar{x}))$ and therefore $\sum_{i=1}^{+\infty} \|x^{(i+1)} - x^{(i)}\| < +\infty$, which implies that the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to some x^*. Since $\bar{x}$ is a limit point of the sequence, it must be $x^* = \bar{x}$. $\square$

Acknowledgments

This work has been partially supported by MIUR under the two projects FIRB - Futuro in Ricerca 2012, contract RBFR12M3AC and PRIN 2012, contract 2012MTE38N, and by the Italian GNCS - INdAM. I. Loris is a Research Associate of the Fonds de la Recherche Scientifique.

References

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

[2] H. Attouch, J. Bolte, and B. F. Svaiter 2013, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.* **137**, 91–129.

[3] J. Barzilai and J. M. Borwein 1988, Two-point step size gradient methods, *IMA J. Numer. Anal.* **8**, 141–148.

[4] S. Bonettini, R. Zanella, and L. Zanni 2009, A scaled gradient projection method for constrained image deblurring, *Inverse Probl.* **25**, 015002.

[5] S. Bonettini and M. Prato 2010, Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm, *Inverse Probl.* **26**, 095001.

[6] S. Bonettini, A. Cornelio, and M. Prato 2013, A new semiblind deconvolution approach for Fourier-based image restoration: an application in astronomy, *SIAM J. Imaging Sci.* **6**, 1736–1757.

[7] S. Bonettini and M. Prato 2014, Accelerated gradient methods for the X-ray imaging of solar flares, *Inverse Probl.* **30**, 055004.

[8] S. Bonettini and M. Prato 2015, New convergence results for the scaled gradient projection method, *Inverse Probl.* **31**, 095008.

[9] S. Bonettini, I. Loris, F. Porta, and M. Prato 2016, Variable metric inexact line–search based methods for nonsmooth optimization, *SIAM J. Optim.* **26**, 891–921.

[10] S. Bonettini, I. Loris, F. Porta, M. Prato, and S. Rebegoldi 2016, On the convergence of variable metric line-search based proximal-gradient method under the Kurdyka-Łojasiewicz inequality, *arxiv*:1605.03791.

[11] E. Chouzenoux, J.-C. Pesquet, and A. Repetti 2014, Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function, *J. Optim. Theory Appl.* **162**, 107–132.

[12] P. L. Combettes and B. C. Vũ 2014, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization* **63**, 17–31.

[13] R. Fletcher 2012, A limited memory steepest descent method, *Math. Program.* **135**, 413–436.

[14] K. Kurdyka 1998, On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* **48**, 769–783.

[15] S. Łojasiewicz 1993, Sur la géométrie semi- et sous-analytique, *Ann. Inst. Fourier* **43**, 1575–1595.

[16] I. Loris, M. Bertero, C. De Mol, R. Zanella, and L. Zanni 2009, Accelerating gradient projection methods for ℓ_1-constrained signal recovery by steplength selection rules, *Appl. Computat. Harmon. A.* **27**, 247–254.

[17] M. Prato, R. Cavicchioli, L. Zanni, P. Boccacci, and M. Bertero 2012, Efficient deconvolution methods for astronomical imaging: algorithms and IDL-GPU codes, *Astron. Astrophys.* **539**, A133.

[18] M. Prato, A. La Camera, S. Bonettini, and M. Bertero 2013, A convergent blind deconvolution method for post-adaptive-optics astronomical imaging, *Inverse Probl.* **29**, 065017.

[19] Y. Yuan 2006, A new stepsize for the steepest descent method, *J. Comput. Math.* **24**, 149–156.

[20] B. Zhou, L. Gao, and Y. H. Dai: Gradient methods with adaptive step-sizes. *Comput. Optim. Appl.* **35**, 69–86.